

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number
WO 02/097059 A2

(51) International Patent Classification⁷: C12N
(21) International Application Number: PCT/US02/17452
(22) International Filing Date: 30 May 2002 (30.05.2002)
(25) Filing Language: English
(26) Publication Language: English
(30) Priority Data:
60/294,758 30 May 2001 (30.05.2001) US
60/366,891 21 March 2002 (21.03.2002) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 60/294,758 (CIP)
Filed on 30 May 2001 (30.05.2001)
US 60/366,891 (CIP)
Filed on 21 March 2002 (21.03.2002)

(71) Applicant (for all designated States except US): CHROMOS MOLECULAR SYSTEMS, INC. [CA/CA]; 8081 Lougheed Highway, Burnaby, B.C. V5A 1W9 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): PERKINS, Edward [US/CA]; 7610 Lawrence Drive, Burnaby, B.C. V5A 1T6 (CA). PEREZ, Carl [US/CA]; 1201-7680 Granville Avenue, Richmond, B.C. V6Y 4B9 (CA). LINDENBAUM, Michael [CA/CA]; 252 Finnigan Street, Coquitlam, B.C. V3K 5J7 (CA). GREENE, Amy [US/CA]; 7610 Lawrence Drive, Burnaby, B.C. V5A 1T6 (CA). LEUNG, Josephine [CA/CA]; 711 Ebert Avenue, Coquitlam, B.C. V3J 7P8 (CA). FLEMING, Elena [CA/CA]; 248 E 18th, North Vancouver, B.C. V7L 2X6 (CA). STEWART, Sandra [CA/CA]; 2618 Oxford Street, Vancouver, B.C. V5K 1N3 (CA). SHELLARD, Joan [CA/CA]; #215-1345 West 15th Avenue, Vancouver, B.C. V6H 3R3 (CA).

(74) Agents: SEIDMAN, Stephanie, L. et al.; Heller Ehrman White & McAuliffe LLP, 7th floor, 4350 La Jolla Village Drive, San Diego, CA 92122-1246 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,

MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: CHROMOSOME-BASED PLATFORMS

(57) Abstract: Artificial chromosomes, including *Aces*, that have been engineered to contain available sites for site-specific, recombination-directed integration of DNA of interest are provided. These artificial chromosomes permit tractable, efficient, rational engineering of the chromosome for a variety of applications.



WO 02/097059 A2

1-

CHROMOSOME-BASED PLATFORMS**RELATED APPLICATIONS**

Benefit of priority to U.S. provisional application Serial No. 60/294,758, filed May 30, 2001, to Perkins, *et al.*, entitled "CHROMOSOME-BASED PLATFORMS" and to U.S. provisional application Serial No. 60/366,891, filed March 21, 2002, to Perkins, *et al.*, entitled

5 "CHROMOSOME-BASED PLATFORMS" is claimed. Where permitted, the subject matter of which are herein incorporated by reference in their entirety.

This application is related to Provisional Application No. 60/294,687, filed May 30, 2001, by CARL PEREZ AND STEVEN

10 FABIJANSKI entitled *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING PLANT ARTIFICIAL CHROMOSOMES* and to U.S. Provisional Application No. 60/296,329, filed June 4, 2001, by CARL PEREZ AND STEVEN FABIJANSKI entitled

15 *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING PLANT ARTIFICIAL CHROMOSOMES*. This application also is related to U.S. Provisional Application No. 60/294,758, filed May 30, 2001, by EDWARD PERKINS *et al.*, entitled *CHROMOSOME-BASED PLATFORMS* and to U.S. Provisional Application No. 60/366,891, filed March 21, 2002, by EDWARD PERKINS *et al.*, entitled

20 *CHROMOSOME-BASED PLATFORMS*. This application is also related to U.S. application Serial Nos. (attorney dkt nos. 24601-419 and 419PC), filed on the same day herewith, entitled *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS OF PREPARING PLANT ARTIFICIAL CHROMOSOMES* to Perez *et al.*.

25 This application is related to U.S. application Serial No. 08/695,191, filed August 7, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND*

-2-

METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES, now U.S. Patent No. 6,025,155. This application is also related to U.S. application Serial No. 08/682,080, filed July 15, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF*

5 *AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*, now U.S. Patent No. 6,077,697. This application is also related U.S. application Serial No. 08/629,822, filed April 10, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING*

10 *ARTIFICIAL CHROMOSOMES* (now abandoned), and is also related to copending U.S. application Serial No. 09/096,648, filed June 12, 1998, by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES* and to U.S. application Serial No.

15 09/835,682, April 10, 1997 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES* (now abandoned). This application is also related to copending U.S. application Serial No. 09/724,726, filed November 28, 2000, U.S. application Serial

20 No. 09/724,872, filed November 28, 2000, U.S. application Serial No. 09/724,693, filed November 28, 2000, U.S. application Serial No. 09/799,462, filed March 5, 2001, U.S. application Serial No. 09/836,911, filed April 17, 2001, and U.S. application Serial No. 10/125,767, filed April 17, 2002, each of which is by GYULA

25 HADLACZKY and ALADAR SZALAY, and is entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*. This application is also related to International PCT application No. WO 97/40183. Where permitted the

-3-

subject matter of each of these provisional applications, international applications, and applications is incorporated by reference in its entirety.

FIELD OF INVENTION

Artificial chromosomes, including *ACes*, that have been engineered
5 to contain available sites for site-specific, recombination-directed
integration of DNA of interest are provided. These artificial chromosomes
permit tractable, efficient, rational engineering of the chromosome.

BACKGROUND

Artificial chromosomes

10 A variety of artificial chromosomes for use in plants and animals,
particularly higher plants and animals are available. In particular, U.S.
Patent Nos. 6,025,155 and 6,077,697 provide heterochromatic artificial
chromosomes designated therein as satellite artificial chromosomes
15 (*SATACs*) and now designated artificial chromosome expression systems
(*ACes*). These chromosomes are prepared by introducing heterologous
DNA into a selected plant or animal cell under conditions that result in
integration into a region of the chromosome that leads to an amplification
event resulting in production of a dicentric chromosome. Subsequent
treatment and growth of cells with dicentric chromosomes, including
20 further amplifications, ultimately results in the artificial chromosomes
provided therein. In order to introduce a desired heterologous gene (or a
plurality of heterologous genes) into the artificial chromosome, the
process is repeated introducing the desired heterologous genes and
nucleic acids in the initial targeting step. This process is time consuming
25 and tedious. Hence, more tractable and efficient methods for introducing
heterologous nucleic acid molecules into artificial chromosomes,
particularly *ACes*, are needed.

-4-

Therefore, it is an object herein to provide engineered artificial chromosomes that permit tractable, efficient and rational engineering of artificial chromosomes.

SUMMARY OF THE INVENTION

5 Provided herein are artificial chromosomes that permit tractable, efficient and rational engineering thereof. In particular, the artificial chromosomes provided herein contain one or a plurality of loci (sites) for site-specific, recombination-directed integration of DNA. Thus, provided herein are platform artificial chromosome expression systems ("platform
10 *ACes*") containing single or multiple site-specific, recombination sites. The artificial chromosomes and *ACes* artificial chromosomes include plant and animal chromosomes. Any recombinase system that effects site-specific recombination is contemplated for use herein.

 In one embodiment, chromosomes, including platform *ACes*, are
15 provided that contain one or more lambda *att* sites designed for recombination-directed integration in the presence of lambda integrase, and that are mutated so that they do not require additional factors. Methods for preparing such chromosomes, vectors for use in the methods, and uses of the resulting chromosomes are also provided.

20 Platform *ACes* containing the recombination site(s) and methods for introducing heterologous nucleic acid into such sites and vectors therefor, are provided.

 Also provided herein is a bacteriophage lambda (λ) integrase site-specific recombination system.

25 Methods using recombinase mediated recombination target gene expression vectors and/or genes for insertion thereof into platform chromosomes and the resulting chromosomes are provided.

 Combinations and kits containing the combinations of vectors encoding a recombinase and integrase and primers for introduction of the

-5-

site recognized thereby are also provided. The kits optionally include instructions for performing site-directed integration or preparation of *ACes* containing such sites.

Also provided herein are mammalian and plant cells comprising the
5 artificial chromosomes and *ACes* described herein. The cells can be nuclear donor cells, stem cells, such as a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell or an embryonic stem cell.

Also provide is a lambda-intR mutain comprising a glutamic acid to arginine change at position 174 of wild-type lambda-integrase3. Also
10 provided are transgenic animals and methods for producing a transgenic animal, comprising introducing a *ACes* into an embryonic cell, such as a stem cell or embryo. The *ACes* can comprise heterologous nucleic acid that encodes a therapeutic product. The transgenic animal can be a fish, insect, reptile, amphibians, arachnid or mammal. In certain embodiments,
15 the *ACes* is introduced by cell fusion, lipid-mediated transfection by a carrier system, microinjection, microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer.

The platform *ACes*, including plant and animal *ACes*, such as MACs, provided herein can be introduced into cells, such as, but not
20 limited to, animal cells, including mammalian cells, and into plant cells. Hence plant cells that contain platform MACs, animal cells that contain platform PACs and other combinations of cells and platform *ACes* are provided.

DESCRIPTION OF FIGURES

25 FIGURE 1 provides a diagram depicting creation of an exemplary *ACes* artificial chromosome prepared using methods detailed in U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183. In this exemplified embodiment, the nucleic acid is targeted to an acrocentric chromosome in an animal or plant, and the

-6-

heterologous nucleic acid includes a sequence-specific recombination site and marker genes.

FIGURE 2 provides a map of pWEPuro9K, which is a targeting vector derived from the vector pWE15 (GenBank Accession # X65279; SEQ ID No. 31). Plasmid pWE15 was modified by replacing the *Sa*I (Klenow filled)/*Sma*I neomycin resistance encoding fragment with the *Pvu*II/*Bam*HI (Klenow filled) puromycin resistance-encoding fragment (isolated from plasmid pPUR, Clontech Laboratories, Inc., Palo Alto, CA; GenBank Accession no. U07648; SEQ ID No. 30) resulting in plasmid pWEPuro. Subsequently a 9 Kb *Not*I fragment from the plasmid pFK161 (see Example 1, see, also Csonka *et al.* (2000) *Journal of Cell Science* 113:3207-32161; and SEQ ID NO: 118), containing a portion of the mouse rDNA region, was cloned into the *Not*I site of pWEPuro resulting in plasmid pWEPuro9K.

FIGURE 3 depicts construction of an ACes platform chromosome with a single recombination site, such as loxP sites or an *att*P or *att*B site. This platform ACes chromosome is an exemplary artificial chromosome with a single recombination site.

FIGURE 4 provides a map of plasmid pSV40-193attPsensePur.

FIGURE 5 depicts a method for formation of a chromosome platform with multiple recombination integration sites, such as *att*P sites.

FIGURE 6 sets forth the sequences of the core region of *att*P, *att*B, *att*L and *att*R (SEQ ID Nos. 33-36).

FIGURE 7 depicts insertional recombination of a vector encoding a marker gene, DsRed and an *att*B site with an artificial chromosome containing an *att*P site.

FIGURE 8 provides a map of plasmid pCXLamIntR (SEQ ID NO: 112), which includes the Lambda integrase (E174R)-encoding nucleic acid.

-7-

FIGURE 9 diagrammatically summarizes the platform technology; marker 1 permits selection of the artificial chromosomes containing the integration site; marker 2, which is promoterless in the target gene expression vector, permits selection of recombinants. Upon

5 recombination with the platform marker 2 is expressed under the control of a promoter resident on the platform.

FIGURE 10 provides the vector map for the plasmid p18attBZEO-5'6XHS4eGFP (SEQ ID NO: 116).

FIGURE 11 provides the vector map for the plasmid p18attBZEO-10 3'6XHS4eGFP (SEQ ID NO: 115).

FIGURE 12 provides the vector map for the plasmid p18attBZEO-(6XHS4)2eGFP (SEQ ID NO: 110).

FIGURES 13 AND 14 depict the integration of a PCR product by site-specific recombination as set forth in Example 8.

15 FIGURE 15 provides the vector map for the plasmid pPACrDNA as set forth in Example 9.A.

DETAILED DESCRIPTION OF THE INVENTION

A. DEFINITIONS

Unless defined otherwise, all technical and scientific terms used
20 herein have the same meaning as is commonly understood by one of skill in the art to which the invention(s) belong. All patents, patent applications, published applications and publications, Genbank sequences, websites and other published materials referred to throughout the entire disclosure herein, unless noted otherwise, are incorporated by reference
25 in their entirety. Where reference is made to a URL or other such identifier or address, it understood that such identifiers can change and particular information on the internet can come and go, but equivalent information can be found by searching the internet. Reference thereto evidences the availability and public dissemination of such information.

-8-

As used herein, nucleic acid refers to single-stranded and/or double-stranded polynucleotides, such as deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), as well as analogs or derivatives of either RNA or DNA. Also included in the term "nucleic acid" are analogs of
5 nucleic acids such as peptide nucleic acid (PNA), phosphorothioate DNA, and other such analogs and derivatives. When referring to probes or primers, optionally labeled, with a detectable label, such as a fluorescent or radiolabel, single-stranded molecules are contemplated. Such molecules are typically of a length such that they are statistically unique
10 and of low copy number (typically less than 5, preferably less than 3) for probing or priming a library. Generally a probe or primer contains at least 14, 16 or 30 contiguous nucleotides of sequence complementary to or identical to a gene of interest. Probes and primers can be 10, 20, 30, 50, 100 or more nucleotides long.

15 As used herein, DNA is meant to include all types and sizes of DNA molecules including cDNA, plasmids and DNA including modified nucleotides and nucleotide analogs.

As used herein, nucleotides include nucleoside mono-, di-, and triphosphates. Nucleotides also include modified-nucleotides, such as,
20 but are not limited to, phosphorothioate nucleotides and deazapurine nucleotides and other nucleotide analogs.

As used herein, heterologous or foreign DNA and RNA are used interchangeably and refer to DNA or RNA that does not occur naturally as part of the genome in which it is present or which is found in a location
25 or locations and/or in amounts in a genome or cell that differ from that in which it occurs in nature. Heterologous nucleic acid is generally not endogenous to the cell into which it is introduced, but has been obtained from another cell or prepared synthetically. Generally, although not necessarily, such nucleic acid encodes RNA and proteins that are not

-9-

normally produced by the cell in which it is expressed. Any DNA or RNA that one of skill in the art would recognize or consider as heterologous or foreign to the cell in which it is expressed is herein encompassed by heterologous DNA. Heterologous DNA and RNA may also encode RNA or
5 proteins that mediate or alter expression of endogenous DNA by affecting transcription, translation, or other regulatable biochemical processes.

Examples of heterologous DNA include, but are not limited to, DNA that encodes a gene product or gene product(s) of interest, introduced for purposes of modification of the endogenous genes or for production of an
10 encoded protein. For example, a heterologous or foreign gene may be isolated from a different species than that of the host genome, or alternatively, may be isolated from the host genome but operably linked to one or more regulatory regions which differ from those found in the unaltered, native gene. Other examples of heterologous DNA include, but
15 are not limited to, DNA that encodes traceable marker proteins, such as a protein that confers traits including, but not limited to, herbicide, insect, or disease resistance; traits, including, but not limited to, oil quality or carbohydrate composition. Antibodies that are encoded by heterologous DNA may be secreted or expressed on the surface of the cell in which the
20 heterologous DNA has been introduced.

As used herein, operative linkage or operative association, or grammatical variations thereof, of heterologous DNA to regulatory and effector sequences of nucleotides, such as promoters, enhancers, transcriptional and translational stop sites, and other signal sequences
25 refers to the relationship between such DNA and such sequences of nucleotides. For example, operative linkage of heterologous DNA to a promoter refers to the physical relationship between the DNA and the promoter such that the transcription of such DNA is initiated from the

-10-

promoter by an RNA polymerase that specifically recognizes, binds to and transcribes the DNA.

In order to optimize expression and/or *in vitro* transcription, it may be necessary to remove, add or alter 5' untranslated portions of the clones to eliminate extra, potential inappropriate alternative translation initiation (*i.e.*, start) codons or other sequences that may interfere with or reduce expression, either at the level of transcription or translation. Alternatively, consensus ribosome binding sites (see, *e.g.*, Kozak (1991) *J. Biol. Chem.* 266:19867-19870) can be inserted immediately 5' of the start codon and may enhance expression.

As used herein, a sequence complementary to at least a portion of an RNA, with reference to antisense oligonucleotides, means a sequence having sufficient complementarity to be able to hybridize with the RNA, preferably under moderate or high stringency conditions, forming a stable duplex. The ability to hybridize depends on the degree of complementarity and the length of the antisense nucleic acid. The longer the hybridizing nucleic acid, the more base mismatches it can contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex.

As used herein, regulatory molecule refers to a polymer of deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) or a polypeptide that is capable of enhancing or inhibiting expression of a gene.

As used herein, recognition sequences are particular sequences of nucleotides that a protein, DNA, or RNA molecule, or combinations thereof, (such as, but not limited to, a restriction endonuclease, a modification methylase and a recombinase) recognizes and binds. For example, a recognition sequence for Cre recombinase (see, *e.g.*, SEQ ID NO:58) is a 34 base pair sequence containing two 13 base pair inverted

-11-

repeats (serving as the recombinase binding sites) flanking an 8 base pair core and designated loxP (see, *e.g.*, Sauer (1994) *Current Opinion in Biotechnology* 5:521-527). Other examples of recognition sequences, include, but are not limited to, attB and attP, attR and attL and others (see, *e.g.*, SEQ ID Nos. 8, 41-56 and 72), that are recognized by the recombinase enzyme Integrase (see, SEQ ID Nos. 37 and 38 for the nucleotide and encoded amino acid sequences of an exemplary lambda phage integrase).

The recombination site designated attB is an approximately 33 base pair sequence containing two 9 base pair core-type Int binding sites and a 7 base pair overlap region; attP (SEQ ID No. 72) is an approximately 240 base pair sequence containing core-type Int binding sites and arm-type Int binding sites as well as sites for auxiliary proteins IHF, FIS, and Xis (see, *e.g.*, Landy (1993) *Current Opinion in Biotechnology* 3:699-707| see, SEQ ID Nos. 8 and 72).

As used herein, a recombinase is an enzyme that catalyzes the exchange of DNA segments at specific recombination sites. An integrase herein refers to a recombinase that is a member of the lambda (λ) integrase family.

As used herein, recombination proteins include excisive proteins, integrative proteins, enzymes, co-factors and associated proteins that are involved in recombination reactions using one or more recombination sites (see, Landy (1993) *Current Opinion in Biotechnology* 3:699-707). The recombination proteins used herein can be delivered to a cell via an expression cassette on an appropriate vector, such as a plasmid, and the like. In other embodiments, the recombination proteins can be delivered to a cell in protein form in the same reaction mixture used to deliver the desired nucleic acid, such as a platform ACes, donor target vectors, and the like.

-12-

As used herein the expression "lox site" means a sequence of nucleotides at which the gene product of the cre gene, referred to herein as Cre, can catalyze a site-specific recombination event. A LoxP site is a 34 base pair nucleotide sequence from bacteriophage P1 (see, 5 *e.g.*, Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398-3402). The LoxP site contains two 13 base pair inverted repeats separated by an 8 base pair spacer region as follows: (SEQ ID NO. 57):

ATAACTTCGTATA ATGTATGC TATACGAAGTTAT

E. coli/DH5Δlac and yeast strain BSY23 transformed with plasmid pBS44 10 carrying two loxP sites connected with a LEU2 gene are available from the American Type Culture Collection (ATCC) under accession numbers ATCC 53254 and ATCC 20773, respectively. The lox sites can be isolated from plasmid pBS44 with restriction enzymes *Eco*RI and *Sal*I, or *Xho*I and *Bam*HI. In addition, a preselected DNA segment can be inserted 15 into pBS44 at either the *Sal*I or *Bam*HI restriction enzyme sites. Other lox sites include, but are not limited to, LoxB, LoxL, LoxC2 and LoxR sites, which are nucleotide sequences isolated from *E. coli* (see, *e.g.*, Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398). Lox sites can also be produced by a variety of synthetic techniques (see, *e.g.*, Ito *et al.* (1982) 20 *Nuc. Acid Res.* 10:1755 and Ogilvie *et al.* (1981) *Science* 270:270).

As used herein, the expression "cre gene" means a sequence of nucleotides that encodes a gene product that effects site-specific recombination of DNA in eukaryotic cells at lox sites. One cre gene can be isolated from bacteriophage P1 (see, *e.g.*, Abremski *et al.* (1983) *Cell* 25 32:1301-1311). *E. coli* DH1 and yeast strain BSY90 transformed with plasmid pBS39 carrying a cre gene isolated from bacteriophage P1 and a GAL1 regulatory nucleotide sequence are available from the American Type Culture Collection (ATCC) under accession numbers ATCC 53255

-13-

and ATCC 20772, respectively. The cre gene can be isolated from plasmid pBS39 with restriction enzymes *XhoI* and *SalI*.

As used herein, site-specific recombination refers to site-specific recombination that is effected between two specific sites on a single
5 nucleic acid molecule or between two different molecules that requires the presence of an exogenous protein, such as an integrase or recombinase.

For example, Cre-lox site-specific recombination can include the following three events:

- 10 a. deletion of a pre-selected DNA segment flanked by lox sites;
- b. inversion of the nucleotide sequence of a pre-selected DNA segment flanked by lox sites; and
- c. reciprocal exchange of DNA segments proximate to
15 lox sites located on different DNA molecules.

This reciprocal exchange of DNA segments can result in an integration event if one or both of the DNA molecules are circular. DNA segment refers to a linear fragment of single- or double-stranded deoxyribonucleic acid (DNA), which can be derived from any source.
20 Since the lox site is an asymmetrical nucleotide sequence, two lox sites on the same DNA molecule can have the same or opposite orientations with respect to each other. Recombination between lox sites in the same orientation results in a deletion of the DNA segment located between the two lox sites and a connection between the resulting ends of the original
25 DNA molecule. The deleted DNA segment forms a circular molecule of DNA. The original DNA molecule and the resulting circular molecule each contain a single lox site. Recombination between lox sites in opposite orientations on the same DNA molecule result in an inversion of the nucleotide sequence of the DNA segment located between the two lox

-14-

sites. In addition, reciprocal exchange of DNA segments proximate to lox sites located on two different DNA molecules can occur. All of these recombination events are catalyzed by the gene product of the cre gene. Thus, the Cre-lox system can be used to specifically delete, invert, or
5 insert DNA. The precise event is controlled by the orientation of lox DNA sequences, in *cis* the lox sequences direct the *Cre* recombinase to either delete (lox sequences in direct orientation) or invert (lox sequences in inverted orientation) DNA flanked by the sequences, while *in trans* the lox sequences can direct a homologous recombination event resulting in the
10 insertion of a recombinant DNA.

As used herein, a chromosome is a nucleic acid molecule, and associated proteins, that is capable of replication and segregation within a cell upon cell division. Typically, a chromosome contains a centromeric region, replication origins, telomeric regions and a region of nucleic acid
15 between the centromeric and telomeric regions.

As used herein, a centromere is any nucleic acid sequence that confers an ability to segregate to daughter cells through cell division. A centromere may confer stable segregation of a nucleic acid sequence, including an artificial chromosome containing the centromere, through
20 mitotic or meiotic divisions, including through both mitotic and meiotic divisions. A particular centromere is not necessarily derived from the same species in which it is introduced, but has the ability to promote DNA segregation in cells of that species.

As used herein, euchromatin and heterochromatin have their
25 recognized meanings. Euchromatin refers to chromatin that stains diffusely and that typically contains genes, and heterochromatin refers to chromatin that remains unusually condensed and that has been thought to be transcriptionally inactive. Highly repetitive DNA sequences (satellite DNA) are usually located in regions of the heterochromatin surrounding

-15-

the centromere (pericentric or pericentromeric heterochromatin). Constitutive heterochromatin refers to heterochromatin that contains the highly repetitive DNA which is constitutively condensed and genetically inactive.

5 As used herein, an acrocentric chromosome refers to a chromosome with arms of unequal length.

 As used herein, endogenous chromosomes refer to genomic chromosomes as found in a cell prior to generation or introduction of an artificial chromosome.

10 As used herein, artificial chromosomes are nucleic acid molecules, typically DNA, that stably replicate and segregate alongside endogenous chromosomes in cells and have the capacity to accommodate and express heterologous genes contained therein. It has the capacity to act as a gene delivery vehicle by accommodating and expressing foreign genes
15 contained therein. A mammalian artificial chromosome (MAC) refers to chromosomes that have an active mammalian centromere(s). Plant artificial chromosomes, insect artificial chromosomes and avian artificial chromosomes refer to chromosomes that include centromeres that function in plant, insect and avian cells, respectively. A human artificial
20 chromosome (HAC) refers to chromosomes that include centromeres that function in human cells. For exemplary artificial chromosomes, see, *e.g.*, U.S. Patent Nos. 6,025,155; 6,077,697; 5,288,625; 5,712,134; 5,695,967; 5,869,294; 5,891,691 and 5,721,118 and published International PCT application Nos, WO 97/40183 and WO 98/08964.
25 Artificial chromosomes include those that are predominantly heterochromatic (formerly referred to as satellite artificial chromosomes (SATACs); see, *e.g.*, U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183), minichromosomes that contain a *de novo* centromere (see, U.S. Patent

-16-

Nos. 5,712,134, 5,891,691 and 5,288,625), artificial chromosomes predominantly made up of repeating nucleic acid units and that contain substantially equivalent amounts of euchromatic and heterochromatic DNA and *in vitro* assembled artificial chromosomes (see, copending U.S. provisional application Serial No. 60/294,687, filed on May 30, 2001).

As used herein, the term "satellite DNA-based artificial chromosome (SATAC)" is interchangeable with the term "artificial chromosome expression system (ACes)". These artificial chromosomes (ACes) include those that are substantially all neutral non-coding sequences (heterochromatin) except for foreign heterologous, typically gene-encoding nucleic acid, that is interspersed within the heterochromatin for the expression therein (see U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183), or that is in a single locus as provided herein. Also included are ACes that may include euchromatin and that result from the process described in U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183 and outlined herein. The delineating structural feature is the presence of repeating units, that are generally predominantly heterochromatin. The precise structure of the ACes will depend upon the structure of the chromosome in which the initial amplification event occurs; all share the common feature of including a defined pattern of repeating units. Generally ACes have more heterochromatin than euchromatin. Foreign nucleic acid molecules (heterologous genes) contained in these artificial chromosome expression systems can include any nucleic acid whose expression is of interest in a particular host cell. Such foreign nucleic acid molecules, include, but are not limited to, nucleic acid that encodes traceable marker proteins (reporter genes), such as fluorescent proteins, such as green, blue or red fluorescent proteins (GFP, BFP and RFP, respectively), other reporter

-17-

genes, such as β -galactosidase and proteins that confer drug resistance, such as a gene encoding hygromycin-resistance. Other examples of heterologous nucleic acid molecules include, but are not limited to, DNA that encodes therapeutically effective substances, such as anti-cancer
5 agents, enzymes and hormones, DNA that encodes other types of proteins, such as antibodies, and DNA that encodes RNA molecules (such as antisense or siRNA molecules) that are not translated into proteins.

As used herein, an artificial chromosome platform, also referred to
10 herein as a "platform ACes" or "ACes platform", refers to an artificial chromosome that has been engineered to include one or more sites for site-specific, recombination-directed integration. In particular, ACes that are so-engineered are provided. Any sites, including but not limited to any described herein, that are suitable for such integration are
15 contemplated. Plant and animal platform ACes are provided. Among the ACes contemplated herein are those that are predominantly heterochromatic (formerly referred to as satellite artificial chromosomes (SATACs); see, *e.g.*, U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183), artificial
20 chromosomes predominantly made up of repeating nucleic acid units and that contain substantially equivalent amounts of euchromatic and heterochromatic DNA resulting from an amplification event depicted in the referenced patent and herein. Included among the ACes for use in generating platforms, are artificial chromosomes that introduce and
25 express heterologous nucleic acids in plants (see, copending U.S. provisional application Serial No. 60/294,687, filed on May 30, 2001). These include artificial chromosomes that have a centromere derived from a plant, and, also, artificial chromosomes that have centromeres that may be derived from other organisms but that function in plants.

-18-

As used herein a "reporter ACes" refers to a an ACes that comprises one or a plurality of reporter constructs, where the reporter construct comprises a reporter gene in operative linkage with a regulatory region responsive to test or known compounds.

5 As used herein, amplification, with reference to DNA, is a process in which segments of DNA are duplicated to yield two or multiple copies of substantially similar or identical or nearly identical DNA segments that are typically joined as substantially tandem or successive repeats or inverted repeats.

10 As used herein, amplification-based artificial chromosomes are artificial chromosomes derived from natural or endogenous chromosomes by virtue of an amplification event, such as one initiated by introduction of heterologous nucleic acid into rDNA in a chromosome. As a result of such an event, chromosomes and fragments thereof exhibiting segmented
15 or repeating patterns arise. Artificial chromosomes can be formed from these chromosomes and fragments. Hence, amplification-based artificial chromosomes refer to engineered chromosomes that exhibit an ordered segmentation that is not observed in naturally occurring chromosomes and that distinguishes them from naturally occurring chromosomes. The
20 segmentation, which can be visualized using a variety of chromosome analysis techniques known to those of skill in the art, correlates with the structure of these artificial chromosomes. In addition to containing one or more centromeres, the amplification-based artificial chromosomes, throughout the region or regions of segmentation are predominantly made
25 up of nucleic acid units also referred to as "amplicons", that is (are) repeated in the region and that have a similar gross structure. Repeats of an amplicon tend to be of similar size and share some common nucleic acid sequences. For example, each repeat of an amplicon may contain a replication site involved in amplification of chromosome segments and/or

-19-

some heterologous nucleic acid that was utilized in the initial production of the artificial chromosome. Typically, the repeating units are substantially similar in nucleic acid composition and may be nearly identical.

- 5 The amplification-based artificial chromosomes differ depending on the chromosomal region that has undergone amplification in the process of artificial chromosome formation. The structures of the resulting chromosomes can vary depending upon the initiating event and/or the conditions under which the heterologous nucleic acid is introduced,
- 10 including modification to the endogenous chromosomes. For example, in some of the artificial chromosomes provided herein, the region or regions of segmentation may be made up predominantly of heterochromatic DNA. In other artificial chromosomes provided herein, the region or regions of segmentation may be made up predominantly of euchromatic DNA or may
- 15 be made up of similar amounts of heterochromatic and euchromatic DNA.

- As used herein an amplicon is a repeated nucleic acid unit. In some of the artificial chromosomes described herein, an amplicon may contain a set of inverted repeats of a megareplicon. A megareplicon represents a higher order replication unit. For example, with reference to
- 20 some of the predominantly heterochromatic artificial chromosomes, the megareplicon can contain a set of tandem DNA blocks (*e.g.*, ~7.5 Mb DNA blocks) each containing satellite DNA flanked by non-satellite DNA or may be made up of substantially rDNA. Contained within the megareplicon is a primary replication site, referred to as the
- 25 megareplicator, which may be involved in organizing and facilitating replication of the pericentric heterochromatin and possibly the centromeres. Within the megareplicon there may be smaller (*e.g.*, 50-300 kb) secondary replicons.

-20-

In artificial chromosomes, such as those provided U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183, the megareplicon is defined by two tandem blocks (~ 7.5 Mb DNA blocks in the chromosomes provided therein). Within each artificial chromosome or among a population thereof, each amplicon has the same gross structure but may contain sequence variations. Such variations will arise as a result of movement of mobile genetic elements, deletions or insertions or mutations that arise, particularly in culture. Such variation does not affect the use of the artificial chromosomes or their overall structure as described herein.

As used herein, amplifiable, when used in reference to a chromosome, particularly the method of generating artificial chromosomes provided herein, refers to a region of a chromosome that is prone to amplification. Amplification typically occurs during replication and other cellular events involving recombination (*e.g.*, DNA repair). Such regions include regions of the chromosome that contain tandem repeats, such as satellite DNA, rDNA, and other such sequences.

As used herein, a dicentric chromosome is a chromosome that contains two centromeres. A multicentric chromosome contains more than two centromeres.

As used herein, a formerly dicentric chromosome is a chromosome that is produced when a dicentric chromosome fragments and acquires new telomeres so that two chromosomes, each having one of the centromeres, are produced. Each of the fragments is a replicable chromosome. If one of the chromosomes undergoes amplification of primarily euchromatic DNA to produce a fully functional chromosome that is predominantly (at least more than 50%) euchromatin, it is a minichromosome. The remaining chromosome is a formerly dicentric chromosome. If one of the chromosomes undergoes amplification,

-21-

whereby heterochromatin (such as, for example, satellite DNA) is amplified and a euchromatic portion (such as, for example, an arm) remains, it is referred to as a sausage chromosome. A chromosome that is substantially all heterochromatin, except for portions of heterologous DNA, is called a predominantly heterochromatic artificial chromosome. Predominantly heterochromatic artificial chromosomes can be produced from other partially heterochromatic artificial chromosomes by culturing the cell containing such chromosomes under conditions such as BrdU treatment that destabilize the chromosome and/or growth under selective conditions so that a predominantly heterochromatic artificial chromosome is produced. For purposes herein, it is understood that the artificial chromosomes may not necessarily be produced in multiple steps, but may appear after the initial introduction of the heterologous DNA. Typically, artificial chromosomes appear after about 5 to about 60, or about 5 to about 55, or about 10 to about 55 or about 25 to about 55 or about 35 to about 55 cell doublings after initiation of artificial chromosome generation, or they may appear after several cycles of growth under selective conditions and BrdU treatment.

As used herein, an artificial chromosome that is predominantly heterochromatic (*i.e.*, containing more heterochromatin than euchromatin, typically more than about 50%, more than about 70%, or more than about 90% heterochromatin) may be produced by introducing nucleic acid molecules into cells, such as, for example, animal or plant cells, and selecting cells that contain a predominantly heterochromatic artificial chromosome. Any nucleic acid may be introduced into cells in such methods of producing the artificial chromosomes. For example, the nucleic acid may contain a selectable marker and/or optionally a sequence that targets nucleic acid to the pericentric, heterochromatic region of a chromosome, such as in the short arm of acrocentric chromosomes and

-22-

nucleolar organizing regions. Targeting sequences include, but are not limited to, lambda phage DNA and rDNA for production of predominantly heterochromatic artificial chromosomes in eukaryotic cells.

After introducing the nucleic acid into cells, a cell containing a
5 predominantly heterochromatic artificial chromosome is selected. Such
cells may be identified using a variety of procedures. For example,
repeating units of heterochromatic DNA of these chromosomes may be
discerned by G-banding and/or fluorescence in situ hybridization (FISH)
techniques. Prior to such analyses, the cells to be analyzed may be
10 enriched with artificial chromosome-containing cells by sorting the cells
on the basis of the presence of a selectable marker, such as a reporter
protein, or by growing (culturing) the cells under selective conditions. It
is also possible, after introduction of nucleic acids into cells, to select
cells that have a multicentric, typically dicentric, chromosome, a formerly
15 multicentric (typically dicentric) chromosome and/or various
heterochromatic structures, such as a megachromosome and a sausage
chromosome, that contain a centromere and are predominantly
heterochromatic and to treat them such that desired artificial
chromosomes are produced. Cells containing a new chromosome are
20 selected. Conditions for generation of a desired structure include, but are
not limited to, further growth under selective conditions, introduction of
additional nucleic acid molecules and/or growth under selective conditions
and treatment with destabilizing agents, and other such methods (see
International PCT application No. WO 97/40183 and U.S. Patent Nos.
25 6,025,155 and 6,077,697).

As used herein, a "selectable marker" is a nucleic acid segment,
generally DNA, that allows one to select for or against a molecule or a
cell that contains it, often under particular conditions. These markers can
encode an activity, such as, but not limited to, production of RNA,

-23-

peptide, or protein, or can provide a binding site for RNA, peptides, proteins, inorganic and organic compounds and compositions. Examples of selectable markers include but are not limited to: (1) nucleic acid segments that encode products that provide resistance against otherwise
5 toxic compounds (e.g., antibiotics); (2) nucleic acid segments that encode products that are otherwise lacking in the recipient cell (e.g., tRNA genes, auxotrophic markers); (3) nucleic acid segments that encode products that suppress the activity of a gene product; (4) nucleic acid segments that encode products that can be identified, such as phenotypic markers,
10 including β -galactosidase, red, blue and/or green fluorescent proteins (FPs), and cell surface proteins; (5) nucleic acid segments that bind products that are otherwise detrimental to cell survival and/or function; (6) nucleic acid segments that otherwise inhibit the activity of any of the nucleic acid segments described in Nos. 1-5 above (e.g., antisense
15 oligonucleotides or siRNA molecules for use in RNA interference); (7) nucleic acid segments that bind products that modify a substrate (e.g. restriction endonucleases); (8) nucleic acid segments that can be used to isolate a desired molecule (e.g. specific protein binding sites); (9) nucleic acid segments that encode a specific nucleotide sequence that can be
20 otherwise non-functional, such as for PCR amplification of subpopulations of molecules; and/or (10) nucleic acid segments, which when absent, directly or indirectly confer sensitivity to particular compounds. Thus, for example, selectable markers include nucleic acids encoding fluorescent proteins, such as green fluorescent proteins, β -galactosidase and other
25 readily detectable proteins, such as chromogenic proteins or proteins capable of being bound by an antibody and FACs sorted. Selectable markers such as these, which are not required for cell survival and/or proliferation in the presence of a selection agent, are also referred to herein as reporter molecules. Other selectable markers, e.g., the

-24-

neomycin phosphotransferase gene, provide for isolation and identification of cells containing them by conferring properties on the cells that make them resistant to an agent, *e.g.*, a drug such as an antibiotic, that inhibits proliferation of cells that do not contain the marker.

- 5 As another example, interference of gene expression by double stranded RNA has been shown in *Caenorhabditis elegans*, plants, *Drosophila*, protozoans and mammals. This method is known as RNA interference (RNAi) and utilizes short, double-stranded RNA molecules (siRNAs). The siRNAs are generally composed of a 19-22bp double-
- 10 stranded RNA stem, a loop region and a 1-4 bp overhang on the 3' end. The reduction of gene expression has been accomplished by direct introduction of the siRNAs into the cell (Harborth J et al., 2001, *J Cell Sci* 114(pt 24):4557-65) as well as the introduction of DNA encoding and expressing the siRNA molecule. The encoded siRNA molecules are under
- 15 the regulation of an RNA polymerase III promoter (see, *e.g.*, Yu et al., 2002, *Proc Natl Acad Sci USA* 99(9):6047-52; Brummelkamp et al., 2002, *Science* 296(5567):550-3; Miyagishi et al., 2002, *Nat Biotechnol* 20(5):497-500; and the like). In certain embodiments, RNAi in mammalian cells may have advantages over other therapeutic methods.
- 20 For example, producing siRNA molecules that block viral genetic activities in infected cells may reduce the effects of the virus. Platform ACes provided herein encoding siRNA molecule(s) are an additional utilization of the platform ACes technology. The platform ACes could be engineered to encode one or more siRNA molecules to create gene "knockdowns". In
- 25 one embodiment, a platform ACes can engineered to encode both the siRNA molecule and a replacement gene. For example, a mouse model or cell culture system could be generated using a platform ACes that has a knockdown of the endogenous mouse gene, by siRNA, and the human gene homolog expressing in place of the mouse gene. The placement of

-25-

siRNA encoding sequences under the regulation of a regulatable or inducible promoter would allow one to temporally and/or spatially control the knockdown effect of the corresponding gene.

As used herein, a reporter gene includes any gene that expresses a detectable gene product, which may be RNA or protein. Generally reporter genes are readily detectable. Examples of reporter genes include, but are not limited to nucleic acid encoding a fluorescent protein, CAT (chloramphenicol acetyl transferase) (Alton *et al.* (1979) *Nature* 282: 864-869) luciferase, and other enzyme detection systems, such as beta-galactosidase; firefly luciferase (deWet *et al.* (1987) *Mol. Cell. Biol.* 7:725-737); bacterial luciferase (Engebrecht and Silverman (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81:4154-4158; Baldwin *et al.* (1984) *Biochemistry* 23:3663-3667); and alkaline phosphatase (Toh *et al.* (1989) *Eur. J. Biochem.* 182:231-238, Hall *et al.* (1983) *J. Mol. Appl. Gen.* 2:101).

As used herein, growth under selective conditions means growth of a cell under conditions that require expression of a selectable marker for survival.

As used herein, an agent that destabilizes a chromosome is any agent known by those skilled in the art to enhance amplification events, and/or mutations. Such agents, which include BrdU, are well known to those skilled in the art.

In order to generate an artificial chromosome containing a particular heterologous nucleic acid of interest, it is possible to include the nucleic acid in the nucleic acid that is being introduced into cells to initiate production of the artificial chromosome. Thus, for example, a nucleic acid can be introduced into a cell along with nucleic acid encoding a selectable marker and/or a nucleic acid that targets to a heterochromatic region of a chromosome. For introducing a heterologous nucleic acid into

-26-

the cell, it can be included in a fragment that includes a selectable marker or as part of a separate nucleic acid fragment and introduced into the cell with a selectable marker during the process of generating the artificial chromosomes. Alternatively, heterologous nucleic acid can be introduced
5 into an artificial chromosome at a later time after the initial generation of the artificial chromosome.

As used herein, the minichromosome refers to a chromosome derived from a multicentric, typically dicentric, chromosome that contains more euchromatic than heterochromatic DNA. For purposes herein, the
10 minichromosome contains a *de novo* centromere (e.g., a neocentromere). In some embodiments, for example, the minichromosome contains a centromere that replicates in animals, e.g., a mammalian centromere or in plants, e.g., a plant centromere.

As used herein, *in vitro* assembled artificial chromosomes or
15 synthetic chromosomes can be either more euchromatic than heterochromatic or more heterochromatic than euchromatic and are produced by joining essential components of a chromosome *in vitro*. These components include at least a centromere, a megareplicator, a telomere and optionally secondary origins of replication.

As used herein, *in vitro* assembled plant or animal artificial
20 chromosomes are produced by joining essential components (at least the centromere, telomere(s), megareplicator and optional secondary origins of replication) that function in plants or animals. In particular embodiments, the megareplicator contains sequences of rDNA, particularly plant or
25 animal rDNA.

As used herein, a plant is a eukaryotic organism that contains, in addition to a nucleus and mitochondria, chloroplasts capable of carrying out photosynthesis. A plant can be unicellular or multicellular and can contain multiple tissues and/or organs. Plants can reproduce sexually or

-27-

asexually and can be perennial or annual in growth. Plants can also be terrestrial or aquatic. The term "plant" includes a whole plant, plant cell, plant protoplast, plant calli, plant seed, plant organ, plant tissue, and other parts of a whole plant.

5 As used herein, stable maintenance of chromosomes occurs when at least about 85%, preferably 90%, more preferably 95%, of the cells retain the chromosome. Stability is measured in the presence of a selective agent. Preferably these chromosomes are also maintained in the absence of a selective agent. Stable chromosomes also retain their
10 structure during cell culturing, suffering no unintended intrachromosomal or interchromosomal rearrangements.

 As used herein, *de novo* with reference to a centromere, refers to generation of an excess centromere in a chromosome as a result of incorporation of a heterologous nucleic acid fragment using the methods
15 herein.

 As used herein, BrdU refers to 5-bromodeoxyuridine, which during replication is inserted in place of thymidine. BrdU is used as a mutagen; it also inhibits condensation of metaphase chromosomes during cell division.

20 As used herein, ribosomal RNA (rRNA) is the specialized RNA that forms part of the structure of a ribosome and participates in the synthesis of proteins. Ribosomal RNA is produced by transcription of genes which, in eukaryotic cells, are present in multiple copies. In human cells, the approximately 250 copies of rRNA genes (i.e., genes which encode rRNA)
25 per haploid genome are spread out in clusters on at least five different chromosomes (chromosomes 13, 14, 15, 21 and 22). In mouse cells, the presence of ribosomal DNA (rDNA, which is DNA containing sequences that encode rRNA) has been verified on at least 11 pairs out of 20 mouse chromosomes (chromosomes 5, 6, 7, 9, 11, 12, 15, 16, 17, 18, and 19)

-28-

(see e.g., Rowe *et al.* (1996) *Mamm. Genome* 7:886-889 and Johnson *et al.* (1993) *Mamm. Genome* 4:49-52). In *Arabidopsis thaliana* the presence of rDNA has been verified on chromosomes 2 and 4 (18S, 5.8S, and 25S rDNA) and on chromosomes 3,4, and 5 (5S rDNA)(see The
5 *Arabidopsis* Genome Initiative (2000) *Nature* 408:796-815). In eukaryotic cells, the multiple copies of the highly conserved rRNA genes are located in a tandemly arranged series of rDNA units, which are generally about 40-45 kb in length and contain a transcribed region and a nontranscribed region known as spacer (i.e., intergenic spacer) DNA
10 which can vary in length and sequence. In the human and mouse, these tandem arrays of rDNA units are located adjacent to the pericentric satellite DNA sequences (heterochromatin). The regions of these chromosomes in which the rDNA is located are referred to as nucleolar organizing regions (NOR) which loop into the nucleolus, the site of
15 ribosome production within the cell nucleus.

As used herein, a megachromosome refers to a chromosome that, except for introduced heterologous DNA, is substantially composed of heterochromatin. Megachromosomes are made up of an array of repeated amplicons that contain two inverted megareplicons bordered by
20 introduced heterologous DNA (see, *e.g.*, Figure 3 of U.S. Patent No. 6,077,697 for a schematic drawing of a megachromosome). For purposes herein, a megachromosome is about 50 to 400 Mb, generally about 250-400 Mb. Shorter variants are also referred to as truncated megachromosomes (about 90 to 120 or 150 Mb), dwarf
25 megachromosomes (~150-200 Mb), and a micro-megachromosome (~50-90 Mb, typically 50-60 Mb). For purposes herein, the term

-29-

megachromosome refers to the overall repeated structure based on an array of repeated chromosomal segments (amplicons) that contain two inverted megareplicons bordered by any inserted heterologous DNA. The size will be specified.

5 As used herein, gene therapy involves the transfer or insertion of nucleic acid molecules into certain cells, which are also referred to as target cells, to produce specific products that are involved in preventing, curing, correcting, controlling or modulating diseases, disorders and deleterious conditions. The nucleic acid is introduced into the selected
10 target cells in a manner such that the nucleic acid is expressed and a product encoded thereby is produced. Alternatively, the nucleic acid may in some manner mediate expression of DNA that encodes a therapeutic product. This product may be a therapeutic compound, which is produced in therapeutically effective amounts or at a therapeutically
15 useful time. It may also encode a product, such as a peptide or RNA, that in some manner mediates, directly or indirectly, expression of a therapeutic product. Expression of the nucleic acid by the target cells within an organism afflicted with a disease or disorder thereby provides for modulation of the disease or disorder. The nucleic acid encoding the
20 therapeutic product may be modified prior to introduction into the cells of the afflicted host in order to enhance or otherwise alter the product or expression thereof.

For use in gene therapy, cells can be transfected *in vitro*, followed by introduction of the transfected cells into an organism. This is often
25 referred to as *ex vivo* gene therapy. Alternatively, the cells can be transfected directly *in vivo* within an organism.

As used herein, therapeutic agents include, but are not limited to, growth factors, antibodies, cytokines, such as tumor necrosis factors and interleukins, and cytotoxic agents and other agents disclosed herein and

-30-

known to those of skill in the art. Such agents include, but are not limited to, tumor necrosis factor, α -interferon, β -interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 (IL-1), interleukin-2 (IL-2), interleukin-6 (IL-6), granulocyte
5 macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), erythropoietin (EPO), pro-coagulants such as tissue factor and tissue factor variants, pro-apoptotic agents such as FAS-ligand, fibroblast growth factors (FGF), nerve growth factor and other
10 growth factors.

As used herein, a therapeutically effective product is a product that is encoded by heterologous DNA that, upon introduction of the DNA into a host, a product is expressed that effectively ameliorates or eliminates the symptoms, manifestations of an inherited or acquired disease or that
15 cures the disease.

As used herein, transgenic plants and animals refer to plants and animals in which heterologous or foreign nucleic acid is expressed or in which the expression of a gene naturally present in the plant or animal has been altered by virtue of introduction of heterologous or foreign
20 nucleic acid.

As used herein, IRES (internal ribosome entry site; see, *e.g.*, SEQ ID No. 27 and nucleotides 2736-3308 SEQ ID No. 28) refers to a region of a nucleic acid molecule, such as an mRNA molecule, that allows internal ribosome entry sufficient to initiate translation, which initiation
25 can be detected in an assay for cap-independent translation (see, *e.g.*, U.S. Patent No. 6,171,821). The presence of an IRES within an mRNA molecule allows cap-independent translation of a linked protein-encoding sequence that otherwise would not be translated.

-31-

Internal ribosome entry site (IRES) elements were first identified in picornaviruses, which elements are considered the paradigm for cap-independent translation. The 5' UTRs of all picornaviruses are long and mediate translational initiation by directly recruiting and binding
5 ribosomes, thereby circumventing the initial cap-binding step. IRES elements are frequently found in viral mRNA, they are rare in non-viral mRNA. Among non-viral mRNA molecules that contain functional IRES elements in their respective 5' UTRs are those encoding immunoglobulin heavy chain binding protein (BiP) (Macejak *et al.* (1991) *Nature*
10 353:90-94); *Drosophila* Antennapedia (Oh *et al.* (1992) *Genes Dev.* 6:1643-1653); D. Ultrabithorax (Ye *et al.* (1997) *Mol. Cell Biol.* 17:1714-21); fibroblast growth factor 2 (Vagner *et al.* (1995) *Mol. Cell Biol.* 15:35-44); initiation factor eIF4G (Gan *et al.* (1998) *J. Biol. Chem.* 273:5006-5012); proto-oncogene c-myc (Nanbru *et al.* (1995) *J. Biol.*
15 *Chem.* 272:32061-32066; Stoneley (1998) *Oncogene* 16:423-428); IRES_H; from the 5'UTR of NRF1 gene (Oumard *et al.* (2000) *Mol. and Cell Biol.*, 20(8):2755-2759); and vascular endothelial growth factor (VEGF) (Stein *et al.* (1998) *Mol. Cell Biol.* 18:3112-9).

As used herein, a promoter, with respect to a region of DNA, refers
20 to a sequence of DNA that contains a sequence of bases that signals RNA polymerase to associate with the DNA and initiate transcription of RNA (such as pol II for mRNA) from a template strand of the DNA. A promoter thus generally regulates transcription of DNA into mRNA. A particular promoter provided herein is the Ferritin heavy chain promoter (excluding
25 the Iron Response Element, located in the 5'UTR), which was joined to the 37bp Fer-1 enhancer element. This promoter is set forth as SEQ ID NO:128. The endogenous Fer-1 enhancer element is located upstream of the Fer-1 promoter (e.g., a Fer-1 oligo was cloned proximal to the core promoter).

-32-

As used herein, isolated, substantially pure nucleic acid, such as, for example, DNA, refers to nucleic acid fragments purified according to standard techniques employed by those skilled in the art, such as that found in Sambrook *et al.* ((2001) Molecular Cloning: A Laboratory
5 Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 3rd edition).

As used herein, expression refers to the transcription and/or translation of nucleic acid. For example, expression can be the transcription of a gene that may be transcribed into an RNA molecule,
10 such as a messenger RNA (mRNA) molecule. Expression may further include translation of an RNA molecule and translated into peptides, polypeptides, or proteins. If the nucleic acid is derived from genomic DNA, expression may, if an appropriate eukaryotic host cell or organism is selected, include splicing of the mRNA. With respect to an antisense
15 construct, expression may refer to the transcription of the antisense DNA.

As used herein, vector or plasmid refers to discrete elements that are used to introduce heterologous nucleic acids into cells for either expression of the heterologous nucleic acid or for replication of the heterologous nucleic acid. Selection and use of such vectors and
20 plasmids are well within the level of skill of the art.

As used herein, transformation/transfection refers to the process by which nucleic acid is introduced into cells. The terms transfection and transformation refer to the taking up of exogenous nucleic acid, e.g., an expression vector, by a host cell whether or not any coding sequences
25 are in fact expressed. Numerous methods of transfection are known to the ordinarily skilled artisan, for example, by *Agrobacterium*-mediated transformation, protoplast transformation (including polyethylene glycol (PEG)-mediated transformation, electroporation, protoplast fusion, and microcell fusion), lipid-mediated delivery, liposomes, electroporation,

-33-

sonoporation, microinjection, particle bombardment and silicon carbide whisker-mediated transformation and combinations thereof (see, *e.g.*, Paszkowski *et al.* (1984) *EMBO J.* 3:2717-2722; Potrykus *et al.* (1985) *Mol. Gen. Genet.* 199:169-177; Reich *et al.* (1986) *Biotechnology* 4:1001-1004; Klein *et al.* (1987) *Nature* 327:70-73; U.S. Patent No. 6,143,949; Paszkowski *et al.* (1989) in *Cell Culture and Somatic Cell Genetics of Plants, Vol. 6, Molecular Biology of Plant Nuclear Genes*, eds. Schell, J and Vasil, L.K. Academic Publishers, San Diego, California, p. 52-68; and Frame *et al.* (1994) *Plant J.* 6:941-948), direct uptake using calcium phosphate (CaPO₄; see, *e.g.*, Wigler *et al.* (1979) *Proc. Natl. Acad. Sci. U.S.A.* 76:1373-1376), polyethylene glycol (PEG)-mediated DNA uptake, lipofection (see, *e.g.*, Strauss (1996) *Meth. Mol. Biol.* 54:307-327), microcell fusion (see, EXAMPLES, see, also Lambert (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88:5907-5911; U.S. Patent No. 5,396,767, Sawford *et al.* (1987) *Somatic Cell Mol. Genet.* 13:279-284; Dhar *et al.* (1984) *Somatic Cell Mol. Genet.* 10:547-559; and McNeill-Killary *et al.* (1995) *Meth. Enzymol.* 254:133-152), lipid-mediated carrier systems (see, *e.g.*, Teifel *et al.* (1995) *Biotechniques* 19:79-80; Albrecht *et al.* (1996) *Ann. Hematol.* 72:73-79; Holmen *et al.* (1995) *In Vitro Cell Dev. Biol. Anim.* 31:347-351; Remy *et al.* (1994) *Bioconjug. Chem.* 5:647-654; Le Bolch *et al.* (1995) *Tetrahedron Lett.* 36:6681-6684; Loeffler *et al.* (1993) *Meth. Enzymol.* 217:599-618) or other suitable method. Methods for delivery of ACes are described in copending U.S. application Serial No. 09/815,979. Successful transfection is generally recognized by detection of the presence of the heterologous nucleic acid within the transfected cell, such as, for example, any visualization of the heterologous nucleic acid or any indication of the operation of a vector within the host cell.

-34-

As used herein, "delivery," which is used interchangeably with "transfection," refers to the process by which exogenous nucleic acid molecules are transferred into a cell such that they are located inside the cell. Delivery of nucleic acids is a distinct process from expression of
5 nucleic acids.

As used herein, injected refers to the microinjection, such as by use of a small syringe, needle, or pipette, for injection of nucleic acid into a cell.

As used herein, substantially homologous DNA refers to DNA that
10 includes a sequence of nucleotides that is sufficiently similar to another such sequence to form stable hybrids, with each other or a reference sequence, under specified conditions.

It is well known to those of skill in this art that nucleic acid fragments with different sequences may, under the same conditions,
15 hybridize detectably to the same "target" nucleic acid. Two nucleic acid fragments hybridize detectably, under stringent conditions over a sufficiently long hybridization period, because one fragment contains a segment of at least about 10, 14 or 16 or more nucleotides in a sequence that is complementary (or nearly complementary) to a substantially
20 contiguous sequence of at least one segment in the other nucleic acid fragment. If the time during which hybridization is allowed to occur is held constant, at a value during which, under preselected stringency conditions, two nucleic acid fragments with complementary base-pairing segments hybridize detectably to each other, departures from exact
25 complementarity can be introduced into the base-pairing segments, and base-pairing will nonetheless occur to an extent sufficient to make hybridization detectable. As the departure from complementarity between the base-pairing segments of two nucleic acids becomes larger, and as

-35-

conditions of the hybridization become more stringent, the probability decreases that the two segments will hybridize detectably to each other.

Two single-stranded nucleic acid segments have "substantially the same sequence", if (a) both form a base-paired duplex with the same
5 segment, and (b) the melting temperatures of the two duplexes in a solution of 0.5 X SSPE differ by less than 10°C. If the segments being compared have the same number of bases, then to have "substantially the same sequence", they will typically differ in their sequences at fewer than 1 base in 10. Methods for determining melting temperatures of
10 nucleic acid duplexes are well known (see, *e.g.*, Meinkoth *et al.* (1984) *Anal. Biochem.* 138:267-284 and references cited therein).

As used herein, a nucleic acid probe is a DNA or RNA fragment that includes a sufficient number of nucleotides to specifically hybridize to DNA or RNA that includes complementary or substantially complementary
15 sequences of nucleotides. A probe may contain any number of nucleotides, from as few as about 10 and as many as hundreds of thousands of nucleotides. The conditions and protocols for such hybridization reactions are well known to those of skill in the art as are the effects of probe size, temperature, degree of mismatch, salt
20 concentration and other parameters on the hybridization reaction. For example, the lower the temperature and higher the salt concentration at which the hybridization reaction is carried out, the greater the degree of mismatch that may be present in the hybrid molecules.

To be used as a hybridization probe, the nucleic acid is generally
25 rendered detectable by labeling it with a detectable moiety or label, such as ^{32}P , ^3H and ^{14}C , or by other means, including chemical labeling, such as by nick-translation in the presence of deoxyuridylate biotinylated at the 5'-position of the uracil moiety. The resulting probe includes the biotinylated uridylate in place of thymidylate residues and can be detected

-36-

(via the biotin moieties) by any of a number of commercially available detection systems based on binding of streptavidin to the biotin. Such commercially available detection systems can be obtained, for example, from Enzo Biochemicals, Inc. (New York, NY). Any other label known to those of skill in the art, including non-radioactive labels, may be used as long as it renders the probes sufficiently detectable, which is a function of the sensitivity of the assay, the time available (for culturing cells, extracting DNA, and hybridization assays), the quantity of DNA or RNA available as a source of the probe, the particular label and the means used to detect the label.

Once sequences with a sufficiently high degree of homology to the probe are identified, they can readily be isolated by standard techniques (see, *e.g.*, Sambrook *et al.* (2001) *Molecular Cloning: A Laboratory Manual*, 3rd Edition, Cold Spring Harbor Laboratory Press).

As used herein, conditions under which DNA molecules form stable hybrids are considered substantially homologous, and a DNA or nucleic acid homolog refers to a nucleic acid that includes a preselected conserved nucleotide sequence, such as a sequence encoding a polypeptide. By the term "substantially homologous" is meant having at least 75%, preferably 80%, preferably at least 90%, most preferably at least 95% homology therewith or a less percentage of homology or identity and conserved biological activity or function.

The terms "homology" and "identity" are often used interchangeably. In this regard, percent homology or identity may be determined, for example, by comparing sequence information using a GAP computer program. The GAP program utilizes the alignment method of Needleman and Wunsch (*J. Mol. Biol.* 48:443 (1970), as revised by Smith and Waterman (*Adv. Appl. Math.* 2:482 (1981)). Briefly, the GAP program defines similarity as the number of aligned symbols (*i.e.*,

-37-

nucleotides or amino acids) which are similar, divided by the total number of symbols in the shorter of the two sequences. The preferred default parameters for the GAP program may include: (1) a unary comparison matrix (containing a value of 1 for identities and 0 for non-identities) and
5 the weighted comparison matrix of Gribskov and Burgess, *Nucl. Acids Res.* 14:6745 (1986), as described by Schwartz and Dayhoff, eds., *ATLAS OF PROTEIN SEQUENCE AND STRUCTURE*, National Biomedical Research Foundation, pp. 353-358 (1979); (2) a penalty of 3.0 for each gap and an additional 0.10 penalty for each symbol in each gap; and (3)
10 no penalty for end gaps.

By sequence identity, the number of conserved amino acids are determined by standard alignment algorithms programs, and are used with default gap penalties established by each supplier. Substantially homologous nucleic acid molecules would hybridize typically at moderate
15 stringency or at high stringency all along the length of the nucleic acid of interest. Preferably the two molecules will hybridize under conditions of high stringency. Also contemplated are nucleic acid molecules that contain degenerate codons in place of codons in the hybridizing nucleic acid molecule.

20 Whether any two nucleic acid molecules have nucleotide sequences that are at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% "identical" can be determined using known computer algorithms such as the "FAST A" program, using for example, the default parameters as in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 (1988).
25 Alternatively the BLAST function of the National Center for Biotechnology Information database may be used to determine relative sequence identity.

In general, sequences are aligned so that the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be

-38-

calculated using published techniques. (See, *e.g.*: *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H. & Lipton, D., *SIAM J Applied Math* 48:1073 (1988)). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H. & Lipton, D., *SIAM J Applied Math* 48:1073 (1988). Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCG program package (Devereux, J., *et al.*, *Nucleic Acids Research* 12(II):387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F., *et al.*, *J Molec Biol* 215:403 (1990)).

Therefore, as used herein, the term "identity" represents a comparison between a test and a reference polypeptide or polynucleotide. For example, a test polypeptide may be defined as any polypeptide that is 90% or more identical to a reference polypeptide.

As used herein, the term at least "90% identical to" refers to percent identities from 90 to 99.99 relative to the reference polypeptides. Identity at a level of 90% or more is indicative of the fact that, assuming for exemplification purposes a test and reference polynucleotide length of

-39-

100 amino acids are compared. No more than 10% (i.e., 10 out of 100) amino acids in the test polypeptide differs from that of the reference polypeptides. Similar comparisons may be made between a test and reference polynucleotides. Such differences may be represented as point mutations randomly distributed over the entire length of an amino acid sequence or they may be clustered in one or more locations of varying length up to the maximum allowable, e.g. 10/100 amino acid difference (approximately 90% identity). Differences are defined as nucleic acid or amino acid substitutions, or deletions.

- 10 As used herein: stringency of hybridization in determining percentage mismatch encompass the following conditions or equivalent conditions thereto:
- 1) high stringency: 0.1 x SSPE or SSC, 0.1% SDS, 65°C
 - 2) medium stringency: 0.2 x SSPE or SSC, 0.1% SDS, 50°C
 - 15 3) low stringency: 1.0 x SSPE or SSC, 0.1% SDS, 50°C
- or any combination of salt and temperature and other reagents that result in selection of the same degree of mismatch or matching. Equivalent conditions refer to conditions that select for substantially the same percentage of mismatch in the resulting hybrids. Additions of ingredients, such as formamide, Ficoll, and Denhardt's solution affect parameters such as the temperature under which the hybridization should be conducted and the rate of the reaction. Thus, hybridization in 5 X SSC, in 20% formamide at 42° C is substantially the same as the conditions recited above hybridization under conditions of low stringency. The recipes for 25 SSPE, SSC and Denhardt's and the preparation of deionized formamide are described, for example, in Sambrook *et al.* (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Chapter 8; see, Sambrook *et al.*, vol. 3, p. B.13, see, also, numerous catalogs that describe commonly used laboratory solutions. It is understood that

-40-

equivalent stringencies may be achieved using alternative buffers, salts and temperatures. As used herein, all assays and procedures, such as hybridization reactions and antibody-antigen reactions, unless otherwise specified, are conducted under conditions recognized by those of skill in the art as standard conditions.

As used herein, conservative amino acid substitutions, such as those set forth in Table 1, are those that do not eliminate biological activity. Suitable conservative substitutions of amino acids are known to those of skill in this art and may be made generally without altering the biological activity of the resulting molecule. Those of skill in this art recognize that, in general, single amino acid substitutions in non-essential regions of a polypeptide do not substantially alter biological activity (see, *e.g.*, Watson *et al. Molecular Biology of the Gene*, 4th Edition, 1987, The Bejacmin/Cummings Pub. co., p.224). Conservative amino acid substitutions are made, for example, in accordance with those set forth in TABLE 1 as follows:

TABLE 1

	Original residue	Conservative substitution
	Ala (A)	Gly; Ser, Abu
20	Arg (R)	Lys, orn
	Asn (N)	Gln; His
	Cys (C)	Ser
	Gln (Q)	Asn
	Glu (E)	Asp
25	Gly (G)	Ala; Pro
	His (H)	Asn; Gln
	Ile (I)	Leu; Val; Met; Nle; Nva
	Leu (L)	Ile; Val; Met; Nle; Nva
	Lys (K)	Arg; Gln; Glu
30	Met (M)	Leu; Tyr; Ile; NLe Val
	Ornithine	Lys; Arg
	Phe (F)	Met; Leu; Tyr
	Ser (S)	Thr
	Thr (T)	Ser
35	Trp (W)	Tyr
	Tyr (Y)	Trp; Phe
	Val (V)	Ile; Leu; Met; Nle; Nva

-41-

Other substitutions are also permissible and may be determined empirically or in accord with known conservative substitutions.

As used herein, the amino acids, which occur in the various amino acid sequences appearing herein, are identified according to their well-known, three-letter or one-letter abbreviations. The nucleotides, which occur in the various DNA fragments, are designated with the standard single-letter designations used routinely in the art.

As used herein, a splice variant refers to a variant produced by differential processing of a primary transcript of genomic DNA that results in more than one type of mRNA.

As used herein, a probe or primer based on a nucleotide sequence includes at least 10, 14, 16, 30 or 100 contiguous nucleotides from the reference nucleic acid molecule.

As used herein, recombinant production by using recombinant DNA methods refers to the use of the well known methods of molecular biology for expressing proteins encoded by cloned DNA.

As used herein, biological activity refers to the *in vivo* activities of a compound or physiological responses that result upon *in vivo* administration of a compound, composition or other mixture. Biological activity, thus, encompasses therapeutic effects and pharmaceutical activity of such compounds, compositions and mixtures. Biological activities may be observed in *in vitro* systems designed to test or use such activities. Thus, for purposes herein the biological activity of a luciferase is its oxygenase activity whereby, upon oxidation of a substrate, light is produced.

-42-

The terms substantially identical or similar varies with the context as understood by those skilled in the relevant art and generally means at least 40, 60, 80, 90, 95 or 98%.

As used herein, substantially identical to a product means
5 sufficiently similar so that the property is sufficiently unchanged so that the substantially identical product can be used in place of the product.

As used herein, substantially pure means sufficiently homogeneous to appear free of readily detectable impurities as determined by standard methods of analysis, such as thin layer chromatography (TLC), gel
10 electrophoresis and high performance liquid chromatography (HPLC), used by those of skill in the art to assess such purity, or sufficiently pure such that further purification would not detectably alter the physical and chemical properties, such as enzymatic and biological activities, of the substance. Methods for purification of the compounds to produce
15 substantially chemically pure compounds are known to those of skill in the art. A substantially chemically pure compound may, however, be a mixture of stereoisomers or isomers. In such instances, further purification might increase the specific activity of the compound.

As used herein, vector (or plasmid) refers to discrete elements that
20 are used to introduce heterologous DNA into cells for either expression or replication thereof. The vectors typically remain episomal, but may be designed to effect integration of a gene or portion thereof into a chromosome of the genome. Also contemplated are vectors that are artificial chromosomes, such as yeast artificial chromosomes and
25 mammalian artificial chromosomes. Selection and use of such vehicles are well known to those of skill in the art. An expression vector includes vectors capable of expressing DNA that is operatively linked with regulatory sequences, such as promoter regions, that are capable of effecting expression of such DNA fragments. Thus, an expression vector

-43-

refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector that, upon introduction into an appropriate host cell, results in expression of the cloned DNA.

Appropriate expression vectors are well known to those of skill in the art and include those that are replicable in eukaryotic cells and/or prokaryotic
5 cells and those that remain episomal or those which integrate into the host cell genome.

As used herein, protein-binding-sequence refers to a protein or peptide sequence that is capable of specific binding to other protein or
10 peptide sequences generally, to a set of protein or peptide sequences or to a particular protein or peptide sequence.

As used herein, a composition refers to any mixture of two or more ingredients. It may be a solution, a suspension, liquid, powder, a paste, aqueous, non-aqueous or any combination thereof.

15 As used herein, a combination refers to any association between two or more items.

As used herein, fluid refers to any composition that can flow. Fluids thus encompass compositions that are in the form of semi-solids, pastes, solutions, aqueous mixtures, gels, lotions, creams and other such
20 compositions.

As used herein, a cellular extract refers to a preparation or fraction that is made from a lysed or disrupted cell.

As used herein, the term "subject" refers to animals, plants, insects, and birds and other phyla, genera and species into which nucleic
25 acid molecules may be introduced. Included are higher organisms, such as mammals, fish, insects and birds, including humans, primates, cattle, pigs, rabbits, goats, sheep, mice, rats, guinea pigs, hamsters, cats, dogs, horses, chicken and others.

-44-

As used herein, flow cytometry refers to processes that use a laser based instrument capable of analyzing and sorting out cells and or chromosomes based on size and fluorescence.

As used herein, the abbreviations for any protective groups, amino
5 acids and other compounds, are, unless indicated otherwise, in accord with their common usage, recognized abbreviations, or the IUPAC-IUB Commission on Biochemical Nomenclature (see, (1972) *Biochem.* 11:942-944).

B. Recombination systems

10 Site-specific recombination systems typically contain three elements: a pair of DNA sequences (the site-specific recombination sequences) and a specific enzyme (the site-specific recombinase). The site-specific recombinase catalyzes a recombination reaction between two site-specific recombination sequences.

15 A number of different site-specific recombinase systems are available and/or known to those of skill in the art, including, but not limited to: the Cre/*lox* recombination system using CRE recombinase (see, *e.g.*, SEQ ID Nos. 58 and 59) from the Escherichia coli phage P1 (see, *e.g.*, Sauer (1993) *Methods in Enzymology* 225:890-900; Sauer *et al.* 20 (1990) *The New Biologist* 2:441-449), Sauer (1994) *Current Opinion in Biotechnology* 5:521-527; Odell *et al.* (1990) *Mol Gen Genet.* 223:369-378; Lasko *et al.* (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:6232-6236; U.S. Patent No. 5,658,772), the FLP/FRT system of yeast using the FLP recombinase (see, SEQ ID Nos. 60 and 61) from the 2 μ episome of 25 *Saccharomyces cerevisiae* (Cox (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80:4223; Falco *et al.* (1982) *Cell* 29:573-584; Golic *et al.* (1989) *Cell* 59:499-509; U.S. Patent No. 5,744,336), the resolvases, including Gin recombinase of phage Mu (Maeser *et al.* (1991) *Mol Gen Genet.* 230:170-176; Klippel, A. *et al* (1993) *EMBO J.* 12:1047-1057; see, *e.g.*,

-45-

SEQ ID Nos. 64-67), Cin, Hin, $\alpha\delta$ Tn3; the Pin recombinase of *E. coli* (see, e.g., SEQ ID Nos. 68 and 69; Enomoto *et al.* (1983) *J. Bacteriol.* 6:663-668), the R/RS system of the pSR1 plasmid of *Zygosaccharomyces rouxii* (Araki *et al.* (1992) *J. Mol. Biol.* 225:25-37; Matsuzaki *et al.* (1990) *J. Bacteriol.* 172: 610-618) and site-specific recombinases from *Kluyveromyces drosophilarius* (Chen *et al.* (1986) *Nucleic Acids Res.* 314:4471-4481) and *Kluyveromyces waltii* (Chen *et al.* (1992) *J. Gen. Microbiol.* 138:337-345). Other systems are known to those of skill in the art (Stark *et al. Trends Genet.* 8:432-439; Utatsu *et al.* (1987) *J. Bacteriol.* 169:5537-5545; see, also, U.S. Patent No. 6,171,861).

Members of the highly related family of site-specific recombinases, the resolvase family, such as $\gamma\delta$, Tn3 resolvase, Hin, Gin, and Cin are also available. Members of this family of recombinases are typically constrained to intramolecular reactions (e.g., inversions and excisions) and can require host-encoded factors. Mutants have been isolated that relieve some of the requirements for host factors (Maeser *et al.* (1991) *Mol. Gen. Genet.* 230:170-176), as well as some of the constraints of intramolecular recombination (see, U.S. Patent No. 6,171,861).

The bacteriophage P1 Cre/lox and the yeast FLP/FRT systems are particularly useful systems for site-specific integration, inversion or excision of heterologous nucleic acid into, and out of, chromosomes, particularly ACes as provided herein. In these systems a recombinase (Cre or FLP) interacts specifically with its respective site-specific recombination sequence (lox or FRT, respectively) to invert or excise the intervening sequences. The sequence for each of these two systems is relatively short (34 bp for lox and 47 bp for FRT).

The FLP/FRT recombinase system has been demonstrated to function efficiently in plant cells (U.S. Patent No. 5,744,386), and, thus, can be used for producing plant artificial chromosome platforms. In

-46-

general, short incomplete FRT sites leads to higher accumulation of excision products than the complete full-length FRT sites. The system catalyzes intra- and intermolecular reactions, and, thus, can be used for DNA excision and integration reactions. The recombination reaction is
5 reversible and this reversibility can compromise the efficiency of the reaction in each direction. Altering the structure of the site-specific recombination sequences is one approach to remedying this situation. The site-specific recombination sequence can be mutated in a manner that the product of the recombination reaction is no longer recognized as
10 a substrate for the reverse reaction, thereby stabilizing the integration or excision event.

In the Cre-lox system, discovered in bacteriophage P1, recombination between loxP sites occurs in the presence of the Cre recombinase (see, *e.g.*, U.S. Patent No. 5,658,772). This system can be
15 used to insert, invert or excise nucleic acid located between two lox sites. Cre can be expressed from a vector. Since the lox site is an asymmetrical nucleotide sequence, lox sites on the same DNA molecule can have the same or opposite orientation with respect to each other. Recombination between lox sites in the same orientation results in a deletion of the DNA
20 segment located between the two lox sites and a connection between the resulting ends of the original DNA molecule. The deleted DNA segment forms a circular molecule of DNA. The original DNA molecule and the resulting circular molecule each contain a single lox site. Recombination between lox sites in opposite orientations on the same DNA molecule
25 result in an inversion of the nucleotide sequence of the DNA segment located between the two lox sites. In addition, reciprocal exchange of DNA segments proximate to lox sites located on two different DNA molecules can occur. All of these recombination events are catalyzed by the product of the Cre coding region.

-47-

Any site-specific recombinase system known to those of skill in the art is contemplated for use herein. It is contemplated that one or a plurality of sites that direct the recombination by the recombinase are introduced into an artificial chromosome to produce platform ACes. The resulting platform ACes are introduced into cells with nucleic acid encoding the cognate recombinase, typically on a vector, and nucleic acid encoding heterologous nucleic acid of interest linked to the appropriate recombination site for insertion into the platform ACes. The recombinase-encoding-nucleic acid may be introduced into the cells on the same vector, or a different vector, encoding the heterologous nucleic acid.

An *E. coli* phage lambda integrase system for ACes platform engineering and for artificial chromosome engineering is provided (Lorbach *et al.* (2000) *J. Mol. Biol* 296:1175-1181). The phage lambda integrase (Landy, A. (1989) *Annu. Rev. Biochem.* 58:913-94) is adapted herein and the cognate *att* sites are provided. Chromosomes, including ACes, engineered to contain one or a plurality of *att* sites are provided, as are vectors encoding a mutant integrase that functions in the absence other factors. Methods using the modified chromosomes and vectors for introduction of heterologous nucleic acid are also provided.

For purposes herein, one or more of the sites (e.g., a single site or a pair of sites) required for recombination are introduced into an artificial chromosome, such as an ACes chromosome. The enzyme for catalyzing site-directed recombination is introduced with the DNA of interest, or separately, or is engineered onto the artificial chromosome under the control of a regulatable promoter.

As described herein, artificial chromosome platforms containing one or multiple recombination sites are provided. The methods and resulting products are exemplified with the lambda phage Att/Int system, but

-48-

similar methods may be used for production of *ACes* platforms with other recombination systems.

The Att/Int system and vectors provided herein are not only intended for engineering *ACes* platforms, but may be used to engineer an Att/Int system into any chromosome. Introduction of att sites into a chromosome will permit engineering of natural chromosomes, such as by permitting targeted integration genes or regulatory regions, and by controlled excision of selected regions. For example, genes encoding a particular trait may be added to a chromosome, such as plant chromosome engineered to contain one or plurality of *att* sites. Such chromosomes may be used for screening DNA to identify genes. Large pieces of DNA can be introduced into cells and the cells screened phenotypically to select those having the desired trait.

C. Platforms

Provided herein are platform artificial chromosomes (platform *ACes*) containing single or multiple site-specific recombination sites. Chromosome-based platform technology permits efficient and tractable engineering and subsequent expression of multiple gene targets. Methods are provided that use DNA vectors and fragments to create platform artificial chromosomes, including animal, particularly mammalian, artificial chromosomes, and plant artificial chromosomes. The artificial chromosomes contain either single or multiple sequence-specific recombination sites suitable for the placement of target gene expression vectors onto the platform chromosome. The engineered chromosome-based platform *ACes* technology is applicable for methods, including cellular and transgenic protein production, transgenic plant and animal production and gene therapy. The platform *ACes* are also useful for producing a library of *ACes* comprising random portions of a given genome (e.g., a mammalian, plant or prokaryotic genome) for genomic

-49-

screening; as well as a library of cells comprising different and/or mutually exclusive *ACes* therein.

Exemplary of artificial chromosome platforms are those based on *ACes*. *ACes* artificial chromosomes are non-viral, self-replicating nucleic acid molecules that function as a natural chromosome, having all the elements required for normal chromosomal replication and maintenance within the cell nucleus. *ACes* artificial chromosomes do not rely on integration into the genome of the cell to be effective, and they are not limited by DNA carrying capacity and as such the therapeutic gene(s) of interest, including regulatory sequences, can be engineered into the *ACes*. In addition, *ACes* are stable *in vitro* and *in vivo* and can provide predictable long-term gene expression. Once engineered and delivered to the appropriate cell or embryo, *ACes* work independently alongside host chromosomes, for *ACes* that are predominantly heterochromatin producing only the products (proteins) from the genes it carries. As provided herein *ACes* are modified by introduction of recombination site(s) to provide a platform for ready introduction of heterologous nucleic acid. The *ACes* platforms can be used for production of transgenic animals and plants; as vectors for genetic therapy; for use as protein production systems; for animal models to identify and target new therapeutics; in cell culture for the development and production of therapeutic proteins; and for a variety of other applications.

1. Generation of artificial chromosomes

Artificial chromosomes may be generated by any method known to those of skill in the art. Of particular interest herein are the *ACes* artificial chromosomes, which contain a repeated unit. Methods for production of *ACes* are described in detail in U.S. Patent Nos. 6,025,155 and 6,077,697, which, as with all patents, applications, publications and other disclosure, are incorporated herein in their entirety.

-50-

Generation of *de novo* ACes.

ACes can be generated by cotransfecting exogenous DNA—such as a mammary tissue specific DNA cassette including the gene sequences for a therapeutic protein, with a rDNA fragment and a drug resistance marker gene into the desired eukaryotic cell, such as plant or animal cells, such as murine cells *in vitro*. DNA with a selectable or detectable marker is introduced, and can be allowed to integrate randomly into pericentric heterochromatin or can be targeted to pericentric heterochromatin, such as that in rDNA gene arrays that reside on acrocentric chromosomes, such as the short arms of acrocentric chromosomes. This integration event activates the “megareplicator” sequence and amplifies the pericentric heterochromatin and the exogenous DNA, and duplicates a centromere. Ensuing breakage of this “dicentric” chromosome can result in the production of daughter cells that contain the substantially-original chromosome and the new artificial chromosome. The resulting ACes contain all the essential elements needed for stability and replication in dividing cells—centromere, origins of replications, and telomeres. ACes have been produced that express marker genes (lacZ, green fluorescent protein, neomycin-resistance, puromycin-resistance, hygromycin-resistance) and genes of interest. Isolated ACes, for example, have been successfully transferred intact to rodent, human, and bovine cells by electroporation, sonoporation, microinjection, and transfection with lipids and dendrimers.

To render the creation of ACes with desired genes more tractable and efficient, “platform” ACes (platform-ACes) can be produced that contain defined DNA sequences for enzyme-mediated homologous DNA recombination, such as by Cre or FLP recombinases (Bouhassira *et al.* (1996) *Blood* 88(supplement 1):190a; Bouhassira *et al.* (1997) *Blood*, 90:3332-3344; Siebler *et al.* (1997) *Biochemistry*: 36:1740-1747;

-51-

Siebler *et al.* (1998) *Biochemistry* 37: 6229-6234; and Bethke *et al.* (1997) *Nucl. Acids Res.* 25:2828-2834), and as exemplified herein the lambda phage integrase. A *lox* site contains two 13 bp inverted repeats to which Cre-recombinase binds and an intervening 8 bp core region.

- 5 Only pairs of sites having identity in the central 6 bp of the core region are proficient for recombination; sites having non-identical core sequences (heterospecific *lox* sites) do not efficiently recombine with each other (Hoess *et al.* (1986) *Nucleic Acids Res.* 14:2287-2300).

10 **Generating acrocentric chromosomes for plant artificial chromosome formation.**

In human and mouse cells *de novo* formation of a satellite DNA based artificial chromosome (SATAC, also referred to as *ACes*) can occur in an acrocentric chromosome where the short arm contains only pericentric heterochromatin, the rDNA array, and telomere sequences.

- 15 Plant species may not have any acrocentric chromosomes with the same physical structure described, but "megareplicator" DNA sequences reside in the plant rDNA arrays, also known as the nucleolar organizing regions (NOR). A structure like those seen in acrocentric mammalian chromosomes can be generated using site-specific recombination between
20 appropriate arms of plant chromosomes.

Approach

- Qin *et al.* ((1994) *Proc. Natl. Acad. Sci. U.S.A.* 91:1706-1710, 1994) describes crossing two *Nicotiana tabacum* transgenic plants. One plant contains a construct encoding a promoterless hygromycin-resistance
25 gene preceded by a *lox* site (*lox-hpt*), the other plant carries a construct containing a cauliflower mosaic virus 35S promoter linked to a *lox* sequence and the *cre* DNA recombinase coding region (35S-*lox-cre*). The constructs were introduced separately by infecting leaf explants with *agrobacterium tumefaciens* which carries the kanamycin-resistance gene

-52-

(Kan^R). The resultant Kan^R transgenic plants were crossed. Plants that carried the appropriate DNA recombination event were identified by hygromycin-resistance.

5 **Modification of the above for generation of ACes**

The Kan^R cultivars are initially screened, such as by FISH, to identify two sets of candidate transgenic plants. One set has one construct integrated in regions adjacent to the pericentric heterochromatin on the short arm of any chromosome. The second set of candidate plants
10 has the other construct integrated in the NOR region of appropriate chromosomes. To obtain reciprocal translocation both sites must be in the same orientation. Therefore a series of crosses are required, Kan^R plants generated, and FISH analyses performed to identify the appropriate "acrocentric" plant chromosome for *de novo* plant ACes formation.

15 2. **Bacteriophage lambda integrase-based site-specific recombination system**

An integral part of the platform technology includes a site-specific recombination system that allows the placement of selected gene targets or genomic fragments onto the platform chromosomes. Any such system
20 may be used. In particular, a method is provided for insertion of additional DNA fragments into the platform chromosome residing in the cell via sequence-specific recombination using the recombinase activity of the bacteriophage lambda integrase. The lambda integrase system is exemplary of the recombination systems contemplated for ACes. Any
25 known recombination system, including any described herein, particularly any that operates without the need for additional factors or that, by virtue of mutation, does not require additional factors, is contemplated.

-53-

As noted the lambda integrase system provided herein can be used with natural chromosomes and artificial chromosomes in addition to ACes. Single or a plurality of recombination sites, which may be the same or different, are introduced into artificial chromosomes to produce artificial chromosome platforms.

3. Creation of bacteriophage lambda integrase site-specific recombination system

The lambda phage-encoded integrase (designated Int) is a prototypical member of the integrase family. Int effects integration and excision of the phage in and out of the *E. coli* genome via recombination between pairs of attachment sites designated *attB/attP* and *attL/attR*. Each *att* site contains two inverted 9 base pair core Int binding sites and a 7 base pair overlap region that is identical in wild-type *att* sites. Each site, except for *attB* contains additional Int binding sites. In flanking regions, there are recognition sequences for accessory DNA binding proteins, such as integration host factor (IHF), factor for inversion stimulation (FIS) and the phage encoded excision protein (XIS). Except for *attB*, Int is a heterobivalent DNA-binding protein and, with assistance from the accessory proteins and negative DNA supercoiling, binds simultaneously to core and arm sites within the same *att* site.

Int, like Cre and FLP, executes an ordered sequential pair of strand exchanges during integrative and excisive recombination. The natural pairs of target sequences for Int, *attB* and *attP* or *attL* and *attR* are located on the same or different DNA molecules resulting in intra or intermolecular recombination, respectively. For example, intramolecular recombination occurs between inversely oriented *attB* and *attP*, or between *attL* and *attR* sequences, respectively, leading to inversion of the intervening DNA segment.

-54-

Like the recombinase systems, such as Cre and FLP, Int directs site-specific recombination. Unlike the other systems, such as Cre and FLP, Int generally requires additional protein factors for integrative and excisive recombination and negative supercoiling for integrative recombination.

5 Hence, the Int system had not been used in eukaryotic targeting systems.

Mutant Int proteins, designated Int-h (E174K) and a derivative thereof Int-h/218(E174K/E218K) do not require accessory proteins to perform intramolecular integrative and excisive recombination in co-

10 296:1175-1181); wild-type Int does not catalyze intramolecular recombination in human cells harboring target sites *attB* and *attP*.

Hence it had been demonstrated that mutant Int can catalyze factor-independent recombination events in human cells.

There has been no demonstration by others that this system can be
15 used for engineering of eukaryotic genomes or chromosomes. Provided herein are chromosomes, including artificial chromosomes, such as but not limited to ACes that contain *att* sites (e.g., platform ACes), and the use of such chromosomes for targeted integration of heterologous DNA into such chromosomes in eukaryotic cells, including animal, such as
20 rodent and human, and plant cells. Mutant Int provided herein is shown to effect site-directed recombination between sites in artificial chromosomes and vectors containing cognate sites.

An additional component of the chromosome-based platform technology is the site-specific integration of target DNA sequences onto
25 the platform. For this the native bacteriophage lambda integrase has been modified to carry out this sequence specific DNA recombination event in eukaryotic cells. The bacteriophage lambda integrase and its cognate DNA substrate *att* is a member of the site-specific recombinase family that also includes the bacteriophage P1 Cre/lox system as well as

-55-

the *Saccharomyces cerevisiae* 2 micron based FLP/FRT system (see, *e.g.*, Landy (1989) *Ann. Rev. Biochem* 58:913-949; Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398-3402; Broach *et al.* (1982) *Cell* 29:227-234).

5 By combining DNA endonuclease and DNA ligase activity these recombinases recognize and catalyze DNA exchanges between sequences flanking the recognition site. During the integration of lambda genome into the *E. coli* (lambda recombination) genome, the phage integrase (INT) in association with accessory proteins catalyzes the DNA exchange
10 between the attP site of the phage genome and the attB site of the bacterial genome resulting in the formation of attL and attR sites (Figure 6). The engineered bacteriophage lambda integrase has been produced herein to carry out an intermolecular DNA recombination event between an incoming DNA molecule (primarily on a vector containing the bacterial
15 attB site) and the chromosome-based platform carrying the lambda attP sequence independent of lambda bacteriophage or bacterial accessory proteins.

In contrast to the bi-directional Cre/lox and FLP/FRT system, the engineered lambda recombination system derived for chromosome-based
20 platform technology is advantageously unidirectional because accessory proteins, which are absent, are required for excision of integrated nucleic acid upon further exposure to the lambda Int recombinase.

4. Creation of platform chromosome containing single or multiple sequence-specific recombination sites

25 **a. Multiple sites**

For the creation of a platform chromosome containing multiple, sequence-specific recombination sites, artificial chromosomes are produced as depicted in Figure 5 and Example 3. As discussed above, artificial chromosomes can be produced using any suitable methodology,

-56-

including those described in U.S. Patent Nos. 5,288,625; 5,712,134; 5,891,691; 6,025,155. Briefly, to prepare artificial chromosomes containing multiple recombination (e.g., integration) sites, nucleic acid (either in the form a one or more plasmids, such as the plasmid pSV40193attP_{sense}PUR set forth in Example 3) is targeted into an amplifiable region of a chromosome, such as the pericentric region of a chromosome. Among such regions are the rDNA gene loci in acrocentric mammalian chromosomes. Hence, targeting nucleic acid for integration into the rDNA region of mammalian acrocentric chromosomes can include the mouse rDNA fragments (for targeting into rodent cell lines) or large human rDNA regions on BAC/PAC vectors (or subclones thereof in standard vectors) for targeting into human acrocentric chromosomes, such as for human gene therapy applications. The targeting nucleic acid generally includes a detectable or selectable marker, such as antibiotic resistance, such as puromycin and hygromycin, a recombination site (such as attP, attB, attL, attR or the like), and/or human selectable markers as required for gene therapy applications. Cells are grown under conditions that result in amplification and ultimately production of ACes artificial chromosomes having multiple recombination (e.g., integration) sites therein. ACes having the desired size are selected for further engineering.

b. Creation of platform chromosome containing a single sequence-specific recombination site

In this method a mammalian platform artificial chromosome is generated containing a single sequence-specific recombination site. In the Example below, this approach is demonstrated using a puromycin resistance marker for selection and a mouse rDNA fragment for targeting into the rDNA locus on mouse acrocentric chromosomes. Other selection markers and targeting DNA sequences as desired and known to those of

-57-

skill in the art can be used. Additional resistance markers include genes conferring resistance to the antibiotics neomycin, blasticidin, hygromycin and zeocin. For applications, such as gene therapy in which potentially immunogenic responses are to be avoided, host, such as human, derived

5 selectable markers or markers detectable with monoclonal antibodies (MAb) followed by fluorescent activated cell sorting (FACS) can be used. Examples in this class include, but are not limited to: human nerve growth factor receptor (detection with MAb); truncated human growth factor receptor (detection with MAb); mutant human dihydrofolate reductase

10 (DHFR; detectable using a fluorescent methotrexate substrate); secreted alkaline phosphatase (SEAP; detectable with fluorescent substrate); thymidylate synthase (TS; confers resistance to fluorodeoxyuridine); human CAD gene (confers resistance to N-phosphonacetyl-L-aspartate (PALA)).

15 To construct a platform artificial chromosome with a single site, an ACes artificial chromosome (or other artificial chromosome of interest) can be produced containing a selectable marker. A single sequence specific recombination site is targeted onto ACes via homologous recombination. For this, DNA sequences containing the site-specific

20 recombination sequence are flanked with DNA sequences homologous to a selected sequence in the chromosome. For example, when using a chromosome containing rDNA or satellite DNA, such DNA can be used as homologous sequences to target the site-specific recombination sequence onto the chromosome. A vector is designed to have these homologous

25 sequences flanking the site-specific recombination site and, after the appropriate restriction enzyme digest to generate free ends of homology to the chromosome, the DNA is transfected into cells harboring the chromosome. After transfection and integration of the site-specific cassette, homologous recombination events onto the platform

-58-

chromosome are subcloned and identified, for example by screening single cell subclones via expression of resistance or a fluorescent marker and PCR analysis. In one embodiment, a platform artificial chromosome, such as a platform *ACes*, that contains a single copy of the recombination site is selected. Examples 2B and 2D exemplify the process, and Figure 3 provides a diagram depicting one method for the creation of a platform mammalian chromosome containing a single sequence-specific recombination site.

5. **Lambda integrase mediated recombination of target gene expression vector onto platform chromosome**

The third component of the chromosome-based platform technology involves the use of target gene expression vectors carrying, for example, genes for gene therapy, genes for transgenic animal or plant production, and those required for cellular protein production of interest. Using lambda integrase mediated site-specific recombination, or any other recombinase-mediated site-specific recombination, the target gene expression vectors are introduced onto the selected chromosome platform. The use of target gene expression vector permits use of the *de novo* generated chromosome-based platforms for a wide range of gene targets. Furthermore, chromosome platforms containing multiple *attP* sites provides the opportunity to incorporate multiple gene targets onto a single platform, thereby providing for expression of multiple gene targets, including the expression of cellular and genetic regulatory genes and the expression of all or parts of metabolic pathways. In addition to expressing small target genes, such as cDNA and hybrid cDNA/artificial intron constructs, the chromosome-based platform can be used for engineering and expressing large genomic fragments carrying target genes along with its endogenous genomic promoter sequences. This is of importance, for example, where the therapy requires precise cell specific

-59-

expression and in instances where expression is best achieved from genomic clones rather than cDNA clones. Figure 9 provides a diagram summarizing one embodiment of the chromosome-based technology.

A feature of the target gene expression vector that is of interest to
5 include is a promoterless marker gene, which as exemplified (see, Figure 9) contains an upstream *attB* site (marker 2 on Figure 9). The nucleic acid encoding the marker is not expressed unless it is placed downstream from a promoter sequence. Using the recombinase technology provided herein, such as the lambda integrase technology (λ INT_{E174R} on figure 8)
10 provided herein, site-specific recombination between the *attB* site on the vector and the promoter-*attP* site (in the "sense" orientation) on the chromosome-based platform results in the expression of marker 2 on the target gene expression vector, thereby providing a positive selection for the lambda INT mediated site-specific recombination event. Site-specific
15 recombination events on the chromosome-based platform versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR. Examples of suitable marker 2 genes, include, but are not limited to, genes that confer resistance to toxic compounds or antibiotics,
20 fluorescence activated cell sorting (FACS) sortable cell surface markers and various fluorescent markers. Examples of these genes include, but are not limited to, human L26a^R (human homolog of *Saccharomyces cerevisiae* CYH⁸ gene), neomycin, puromycin, blasticidin, CD24 (see, e.g., US Patents 5,804,177 and 6,074,836), truncated CD4, truncated low
25 affinity nerve growth factor receptor (LNGFR), truncated LDL receptor, truncated human growth hormone receptor, GFP, RFP, BFP.

The target gene expression vectors contain a gene (target gene) for expression from the chromosome platform. The target gene can be expressed using various constitutive or regulated promoter systems

-60-

- across various mammalian species. For the expression of multiple target genes within the same target gene expression vector, the expression of the multiple targets can be coordinately regulated via viral-based or human internal ribosome entry site (IRES) elements (see, *e.g.*, Jackson *et al.* (1990) *Trends Biochem Sci.* 15: 477-83; Oumard *et al.* (2000) *Mol. Cell. Biol.* 20: 2755-2759). Furthermore, using IRES type elements linked to a downstream fluorescent marker, *e.g.*, green, red or blue fluorescent proteins (GFP, RFP, BFP) allows for the identification of high expressing clones from the integrated target gene expression vector.
- 10 In certain embodiments described herein, the promoterless marker can be transcriptionally downstream of the heterologous nucleic acid, wherein the heterologous nucleic acid encodes a heterologous protein, and wherein the expression level of the selectable marker is transcriptionally linked to the expression level of the heterologous protein.
- 15 In addition, the selectable marker and the heterologous nucleic acid can be transcriptionally linked by the presence of a IRES between them. As set forth herein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic or fluorescent.
- 20 Expression from the target gene expression vector integrated onto the chromosome-based platform can be further enhanced using genomic insulator/boundary elements. The incorporation of insulator sequences into the target gene expression vector helps define boundaries in chromatin structure and thus minimizes influence of chromatin position effects/gene silencing on the expression of the target gene (Bell *et al.* (1999) *Current Opinion in Genetics and Development* 9:191-198; Emery *et al.* (2000) *Proc. Natl. Acad. Sci. U.S.A.* 97:9150-9155). Examples of insulator elements that can be included onto target gene expression vector in order to optimize expression include, but are not limited to:
- 25

-61-

- 1) chicken β -globin HS4 element (Prioleau *et al.* (1999) *EMBO J* 18: 4035-4048);
- 2) matrix attachment regions (MAR; see, *e.g.*, Ramakrishnan *et al.* (2000) *Mol Cell. Biol.* 20:868-877);
- 5 3) scaffold attachment regions (SAR; see, *e.g.*, Auten *et al.* (1999) *Human Gene Therapy* 10:1389-1399); and
- 4) universal chromatin opening elements (UCOE; WO/0005393 and WO/0224930)

The copy number of the target gene can be controlled by sequentially adding multiple target gene expression vectors containing the target gene onto multiple integration sites on the chromosome platform. Likewise, the copy number of the target gene can be controlled within an individual target gene expression vector by the addition of DNA sequences that promote gene amplification. For example, gene amplification can be induced utilizing the dihydrofolate reductase (DHFR) minigene with subsequent selection with methotrexate (see, *e.g.*, Schimke (1984) *Cell* 37:705-713) or amplification promoting sequences from the rDNA locus (see, *e.g.*, Wegner *et al.* (1989) *Nucl. Acids Res.* 17: 9909-9932).

20 6. Platforms with other recombinase system sites

A "double *lox*" targeting strategy mediated by Cre-recombinase (Bethke *et al.* (1997) *Nucl. Acids Res.* 25:2828-2834) can be used. This strategy employs a pair of heterospecific *lox* sites—*loxA* and *loxB*, which differ by one nucleotide in the 8 bp spacer region. Both sites are engineered into the artificial chromosome and also onto the targeting DNA vector. This allows for a direct site-specific insertion of a commercially relevant gene or genes by a Cre-catalyzed double crossover event. In essence a platform ACes is engineered with a hygromycin-resistance gene flanked by the double *lox* sites generating *lox-ACes*, which is maintained

-62-

in the thymidine kinase deficient cell, LMtk(-). The gene of interest, for example, for testing purposes, the green fluorescence protein gene, GFP and a HSV thymidine kinase gene (tk) marker, are engineered between the appropriate *lox* sites of the targeting vector. The vector DNA is

5 cotransfected with plasmid pBS185 (Life Technologies) encoding the *Cre* recombinase gene into mammalian cells maintaining the dual-*lox* artificial chromosome. Transient expression of the *Cre* recombinase catalyzes the site-specific insertion of the gene and the tk-gene onto the artificial chromosome. The transfected cells are grown in HAT medium that

10 selects for only those cells that have integrated and expressed the thymidine kinase gene. The HAT^R colonies are screened by PCR analyses to identify artificial chromosomes with the desired insertion.

To generate the *lox-ACes*, Lambda-Hyg^R-*lox* DNA is transfected into the LMtk(-) cell line harboring the precursor *ACes*. Hygromycin-

15 resistant colonies are analyzed by FISH and Southern blotting for the presence of a single copy insert on the *ACes*.

To demonstrate the gene replacement technology, cell lines containing candidate *lox-ACes* are cotransfected with pTK-GFP-*lox* and pBS185 (encoding the *Cre* recombinase gene) DNA. After transfection,

20 transient expression of plasmid pBS185 will provide sufficient burst of *Cre* recombinase activity to catalyze DNA recombination at the *lox* sites. Thus, a double crossover event between the *ACes* target and the exogenous targeting plasmid carrying the *loxA* and *loxB* permits the simple replacement of the hygromycin-resistance gene on the *lox-ACes*

25 for the tk-GFP cassette from the targeting plasmid, with no integration of vector DNA. Transfected cells are grown in HAT-media to select for tk-expression. Correct targeting will result in the generation of HAT^R, hygromycin sensitive, and green fluorescent cells. The desired integration event is verified by Southern and PCR analyses. Specific PCR primer sets

-63-

are used to amplify DNA sequences flanking the individual *loxA* and *loxB* sites on the *lox-ACes* before and after homologous recombination.

D. Exemplary applications of the Platform *ACes*

Platform *ACes* are applicable and tractable for different/optimized
5 cell lines. Those that include a fluorescent marker, for example, can be
purified and isolated using fluorescent activated cell sorting (FACS), and
subsequently delivered to a target cell. Those with selectable markers
provide for efficient selection and provide a growth advantage. Platform
ACes allow multiple payload delivery of donor target vectors via a
10 positive-selection site-specific, recombination system, and they allow for
the inclusion of additional genetic factors that improve protein production
and protein quality.

The construction and use of the platform *ACes* as provided for
each application may be similarly applied to other applications. Particular
15 descriptions are for exemplification.

1. Cellular Protein Production Platform *ACes* (CPP *ACes*)

As described herein, *ACes* can be produced from acrocentric
chromosomes in rodent (mouse, hamster) cell lines via megareplicator
induced amplification of heterochromatin/rDNA sequences. Such *ACes*
20 are ideal for cellular protein production as well as other applications
described herein and known to those of skill in the art. *ACes* platforms
that contain a plurality of recombination sites are particularly suitable for
engineering as cellular protein production systems.

In one embodiment, CPP *ACes* involve a two-component system:
25 the platform chromosome containing multiple engineering sites and the
donor target vector containing a platform-specific recombination site with
designed expression cassettes (see Figure 9).

The platform *ACes* can be produced from any artificial
chromosome, particularly the amplification-based artificial chromosomes.

-64-

For exemplification, they are produced from rodent artificial chromosomes produced from acrocentric chromosomes using the technology of U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183, in which nucleic acid is targeted to the pericentric heterochromatic, and, particularly into rDNA to initiate the replication event(s). The ACes can be produced directly in the chosen cellular protein production cell lines, such as, but not limited to, CHO cells, hybridomas, plant cells, plant tissues, plant protoplasts, stem cells and plant calli.

10 a. **Platform Construction**

In the exemplary embodiment, the initial *de novo* platform construction requires co-transfecting with excess targeting DNA, such as, rDNA or lambda DNA without an *attP* region, and an engineered selectable marker. The engineered selectable marker should contain promoter, generally a constitutive promoter, such as human, viral, i.e., adenovirus or SV40 promoter, including the human ferritin heavy chain promoter (SEQ ID NO:128), SV40 and EF1 α promoters, to control expression of a marker gene that provides a selective growth advantage to the cell. An example of such a marker gene is the *E. coli hisD* gene (encoding histidinol dehydrogenase) which is homologous and analogous to the *S. typhimurium hisD* a dominant marker selection system for mammalian cells previously described (see, Hartman *et al.* (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85:8047-8051). Since histidine is an essential amino acid in mammals and a nutritional requirement in cell culture, the *E. coli hisD* gene can be used to select for histidine prototrophy in defined media. Furthermore more stringent selection can be placed on the cells by including histidinol in the medium. Histidinol is itself permeable and toxic to cells. The *hisD* provides a means of detoxification.

-65-

Placed between the promoter and the marker gene is the bacteriophage lambda *attP* site to use the bacteriophage lambda integrase dependent site-specific recombination system (described herein). The insertion of an *attP* site downstream of a promoter element provide forward selection of site-specific recombination events onto the platform *ACes*.

b. Donor Target Vector Construction

A second component of the CPP platform *ACes* system involves the construction of donor target vectors containing a gene product(s) of interest for the CPP platform *ACes*. Individual donor target vectors can be designed for each gene product to be expressed thus enabling maximum usage of a *de novo* constructed platform *ACes*, so that one or a few CPP platform *ACes* will be required for many gene targets.

A key feature of the donor vector target is the *promoterless* marker gene containing an upstream *attB* site (marker 2 on figure 9). Normally the marker would not be expressed unless it is placed downstream of a promoter sequence. As discussed above, using the lambda integrase technology (λ INT_{E174R} on Figure 8 and Figure 9), site-specific recombination between the *attB* site on the vector and the promoter-*attP* site on the CPP platform *ACes* result in the expression of the donor target vector marker providing positive selection for the site-specific event. Site-specific recombination events on the CPP *ACes* versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR. In addition, since the lambda integrase reaction is unidirectional, i.e. excision reaction is not possible, a number of unique targets can be loaded onto the CPP platform *ACes* limited only by the number of markers available.

-66-

Additional features of the donor target vector include gene target expression cassettes flanked by either chromatin insulator regions, matrix attachment regions (MAR) or scaffold attachment regions (SAR). The use of these regions will provide a more "open" chromatin environment for gene expression and help alleviate silencing. An example of such a cassette for expressing a monoclonal antibody is described. For this purpose, a strong constitutive promoter, e.g. chicken β -actin or RNA Poll, is used to drive the expression of the heavy and light chain open reading frames. The heavy and light chain sequences flank a nonattenuated human IRES (IRES_H; from the 5'UTR of NRF1 gene; see Oumard et al., 2000, *Mol. and Cell Biol.*, 20(8):2755-2759) element thereby coordinating transcription of both heavy and light chain sequence. Distal to the light chain open reading frame resides an additional viral encoded IRES (IRES_V modified ECMV internal ribosomal entry site (IRES)) element attenuating the expression of the fluorescent marker gene hrGFP from *Renilla* (Stratagene). By linking the hrGFP with an attenuated IRES, the heavy and light chains along with the hrGFP are monocistronic. Thus, the identification of hrGFP fluorescing cells will provide a means to detect protein producing cells. In addition, high producing cell lines can be identified and isolated by FACS thereby decreasing the time frame in finding high expressers. Functional monoclonal antibody will be confirmed by ELISA.

c. Additional components in cellular protein production platform ACes (CPP ACes)

In addition to the aforementioned CPP ACes system, other genetic factors can be included to enhance the yield and quality of the expressed protein. Again to provide maximum flexibility, these additional factors can be inserted onto the CPP platform ACes by λ INTE174R dependent site-specific recombination. Other factors that could be used with a CPP

-67-

Platform ACes include for example, adenovirus E1a transactivation system which upregulates both cellular and viral promoters (see, e.g., Svensson and Akusjarvi (1984) EMBO 3:789-794; and US patents 5,866,359; 4,775,630 and 4,920,211).

5 d. **Targets for CHO-ACes engineering to enhance cell growth, such as CHO cell growth and protein production/ quality**

 If adding these additional factors onto the CPP ACes is not prudent or desired, the host cell, CHO cells, can be engineered to express these
10 factors (see, below, targets for CHO-ACes engineering to enhance CHO cell growth and protein production/quality). Additional factors to consider including are addition of insulin or IGF-1 to sustain viability; human sialyltransferases or related factors to produce more human-like glycoproteins; expression of factors to decrease ammonium accumulation
15 during cell growth; expression of factors to inhibit apoptosis; expression of factors to improve protein secretion and protein folding; and expression of factors to permit serum-free transfection and selection.

 1) **Addition of insulin or IGF-1 to sustain viability**

20 Stimulatory factors and/or their receptors are expressed to set up an autocrine loop, to improve cell growth, such as CHO cell growth. Two exemplary candidates are insulin and IGF-1 (see, Biotechnol Prog 2000 Sep;16(5):693-7). Insulin is the most commonly used growth factor for sustaining cell growth and viability in serum-free Chinese hamster ovary
25 (CHO) cell cultures. Insulin and IGF-1 analog (LongR(3) serve as growth and viability factors for CHO cells.

 CHO cells were modified to produce higher levels of essential nutrients and factors. A serum-free (SF) medium for dihydrofolate reductase-deficient Chinese hamster ovary cells (DG44 cells) was
30 prepared. Chinese hamster ovary cells (DG44 cells), which are normally

-68-

maintained in 10% serum medium, were gradually weaned to 0.5% serum medium to increase the probability of successful growth in SF medium (see, Kim *et al.* (199) *In Vitro Cell Dev Biol Anim* 35(4):178-82). A SF medium (SF-DG44) was formulated by supplementing the basal medium with these components; basal medium was prepared by supplementing Dulbecco's modified Eagle's medium and Ham's nutrient mixture F12 with hypoxanthine (10 mg/l) and thymidine (10 mg/l). Development of a SF medium for DG44 cells was facilitated using a Plackett-Burman design technique and weaning of cells.

10

2) Human sialyltransferases or related factors to produce more human-like glycoproteins

CHO cells have been modified by increasing their ability to process protein via addition of complex carbohydrates. This has been achieved by overexpression of relevant processing enzymes, or in some cases, reducing expression of relevant enzymes (see, Bragonzi *et al.* (2000) *Biochim Biophys Acta* 1474(3):273-282; see, also Weikert *et al.* (1999) *Nature biotech.* 17:1116-11121; Ferrari J *et al.* (1998) *Biotechnol Bioeng* 60(5):589-95). A CHO cell line expressing alpha2,6-sialyltransferase was developed for the production of human-like sialylated recombinant glycoproteins. The sialylation defect of CHO cells can be corrected by transfecting the alpha2,6-sialyltransferase (alpha2,6-ST) cDNA into the cells. Glycoproteins produced by such CHO cells display alpha2,6- and alpha2,3-linked terminal sialic acid residues, similar to human glycoproteins.

As another example for improving the production of human-like sialylated recombinant glycoproteins, a CHO cell line has been developed that constitutively expresses sialidase antisense RNA (see, Ferrari J *et al.* (1998) *Biotechnol Bioeng* 60(5):589-95). Several antisense expression

30

-69-

vectors were prepared using different regions of the sialidase gene. Co-transfection of the antisense constructs with a vector conferring puromycin resistance gave rise to over 40 puromycin resistant clones that were screened for sialidase activity. A 5' 474 bp coding segment of the sialidase cDNA, in the inverted orientation in an SV 40-based expression vector, gave maximal reduction of the sialidase activity to about 40% wild-type values.

Oligosaccharide biosynthesis pathways in mammalian cells have been engineered for generation of recombinant glycoproteins (see, *e.g.*, Sburlati (1998) *Biotechnol Prog* 14(2):189-92), which describes a Chinese hamster ovary (CHO) cell line capable of producing bisected oligosaccharides on glycoproteins. This cell line was created by overexpression of a recombinant N-acetylglucosaminyltransferase III (GnT-III) (see, also, Prati *et al.* (1998) *Biotechnol Bioeng* 59(4):445-50, which describes antisense strategies for glycosylation engineering of CHO cells).

3) Expression of factors to decrease ammonium accumulation during cell growth

Excess ammonium, which is a by-product of CHO cell metabolism can have detrimental effects on cell growth and protein quality (see, Yang *et al.* (2000) *Biotechnol Bioeng* 68(4):370-80). To solve this problem ammonium levels were modified by overexpressing carbamoyl phosphate synthetase I and ornithine transcarbamoylase or glutamine synthetase in CHO cells. Such modification resulted in reduced ammonium levels observed and an increase in the growth rate (see Kim *et al.* (2000) *J Biotechnol* 81(2-3):129-40; and Enosawa *et al.* (1997) *Cell Transplant* 6(5):537-40).

4) Expression of factors to improve protein secretion and protein folding

-70-

Overexpression of relevant enzymes can be engineered into the ACes to improve protein secretion and folding.

5) Expression of factors to permit serum-free transfection and selection

- 5 It is advantageous to have the ability to convert CHO cells in suspension growing in serum free medium to adherence without having to resort to serum addition. Laminin or fibronectin addition is sufficient to make cells adherent (see, *e.g.*, Zaworski *et al.* (1993) *Biotechniques* 15(5):863-6) so that expressing either of these genes in CHO cells under
10 an inducible promoter should allow for reversible shift to adherence without requiring serum addition.

2. Platform ACes and Gene Therapy

- The platform ACes provided herein are contemplated for use in mammalian gene therapy, particularly human gene therapy. Human ACes
15 can be derived from human acrocentric chromosomes from human host cells, in which the amplified sequences are heterochromatic and/or human rDNA. Different platform ACes applicable for different tissue cell types are provided. The ACes for gene therapy can contain a single copy of a therapeutic gene inserted into a defined location on platform ACes.
20 Therapeutic genes include genomic clones, cDNA, hybrid genes and other combinations of sequences. Preferred selectable markers are those from the mammalian host, such as human derived factors so that they are non-immunogenic, non-toxic and allow for efficient selection, such as by FACS and/or drug resistance.
25 Platform ACes, useful for gene therapy and other applications, as noted herein, can be generated by megareplicator dependent amplification, such as by the methods in U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183. In one embodiment, human ACes are produced using

-71-

human rDNA constructs that target rDNA arrays on human acrocentric chromosomes and induce the megareplicator in human cells, particularly in primary cell lines (with sufficient number of doublings to form the ACes) or stem cells (such as hematopoietic stem cells, mesenchymal stem cells, adult stem cells or embryonic stem cells) to avoid the introduction of potentially harmful rearranged DNA sequences present in many transformed cell lines. Megareplicator induced ACes formation can result in multiple copies of targeting DNA/selectable markers in each amplification block on both chromosomal arms of the platform ACes.

10 In view of the considerations regarding immunogenicity and toxicity, the production of human platform ACes for gene therapy applications employs a two component system analogous to the platform ACes designed for cellular protein production (CPP platform ACes). The system includes a platform chromosome of entirely human DNA origin
15 containing multiple engineering sites and a gene target vector carrying the therapeutic gene of interest.

a. Platform Construction

The initial *de novo* construction of the platform chromosome employs the co-transfection of excess targeting DNA and a selectable
20 marker. In one embodiment, the DNA is targeted to the rDNA arrays on the human acrocentric chromosomes (chromosomes 13, 14, 15, 21 and 22). For example, two large human rDNA containing PAC clones 18714 and 18720 and the human PAC clone 558F8 are used for targeting (Genome Research (ML) now Incyte, BACPAC Resources, 747 52nd
25 Street, Oakland CA). The mouse rDNA clone pFK161 (SEQ ID NO: 118), which was used to make the human SATAC from the 94-3 hamster/human hybrid cell line (see, *e.g.*, published International PCT application No. WO 97/40183 and Csonka, *et al*, *Journal of Cell Science*

-72-

113:3207-32161 and Example 1 for a description of pFK161) can also be used.

For animal applications, selectable markers should be non-immunogenic in the animal, such as a human, and include, but are not limited to: human nerve growth factor receptor (detected with a MAb, such as described in US patent 6,365,373); truncated human growth factor receptor (detected with MAb), mutant human dihydrofolate reductase (DHFR; fluorescent MTX substrate available); secreted alkaline phosphatase (SEAP; fluorescent substrate available); human thymidylate synthase (TS; confers resistance to anti-cancer agent fluorodeoxyuridine); human glutathione S-transferase alpha (GSTA1; conjugates glutathione to the stem cell selective alkylator busulfan; chemoprotective selectable marker in CD34+ cells); CD24 cell surface antigen in hematopoietic stem cells; human CAD gene to confer resistance to N-phosphonacetyl-L-aspartate (PALA); human multi-drug resistance-1 (MDR-1; P-glycoprotein surface protein selectable by increased drug resistance or enriched by FACS); human CD25 (IL-2 α ; detectable by Mab-FITC); Methylguanine-DNA methyltransferase (MGMT; selectable by carmustine); and Cytidine deaminase (CD; selectable by Ara-C).

Since megareplicator induced amplification generates multiple copies of the selectable marker, a second consideration for the selection of the human marker is the resulting dose of the expressed marker after ACes formation. High level of expression of certain markers may be detrimental to the cell and/or result in autoimmunity. One method to decrease the dose of the marker protein is by shortening its half-life, such as via the fusion of the well-conserved human ubiquitin tag (a 76 amino acid sequence) thus leading to increased turnover of the selectable marker. This has been used successfully for a number of reporter

-73-

systems including DHFR (see, *e.g.*, Stack *et al.* (2000) *Nature Biotechnology* 18:1298-1302 and references cited therein).

Using the ubiquitin tagged protein, a human selectable marker system analogous to the CPP ACes described herein is constructed.

- 5 Briefly, a tagged selectable marker, such as for example one of those described herein, is cloned downstream of an *attP* site and expressed from a human promoter. Exemplary promoters contemplated for use herein include, but are not limited to, the human ferritin heavy chain promoter (SEQ ID NO:128); RNA PolI; EF1 α ; TR; glyceraldehyde-3-
10 phosphate dehydrogenase core promoter (GAP); a GAP core promoter including a proximal insulin inducible element the intervening GAP sequence; phosphofructokinase promoter; and phosphoglycerate kinase promoter. Also contemplated herein is an aldolase A promoter H1 & H2 (representing closely spaced transcriptional start sites) along with the
15 proximal H enhancer. There are 4 promoters (*e.g.*, transcriptional start sites) for this gene, each having different regulatory and tissue activity. The H (most proximal 2) promoters are ubiquitously expressed off the H enhancer. This resulting marker can then be co-transfected along with excess human rDNA targeting sequence into the host cells. An important
20 criteria for the selection of the

- recipient cells is sufficient number of cell doublings for the formation and detection of ACes. Accordingly, the co-transfections should be attempted in human primary cells that can be cultured for long periods of
25 time, such as for example, stem cells (*e.g.*, hematopoietic, mesenchymal, adult or embryonic stem cells), or the like. Additional cell types, include, but are not limited to: single gene transfected cells exhibiting increased life-span; over-expressing c-myc cells, *e.g.* MSU1.1 (Morgan *et al.*, 1991, *Exp. Cell Res.*, Nov;197(1):125-136); over-expressing telomerase lines,

-74-

such as TERT cells; SV40 large T-antigen transfected lines; tumor cell lines, such as HT1080; and hybrid human cell lines, such as the 94-3 hamster/human hybrid cell line.

b. Gene Target Vector

5 The second component of the GT platform *ACes* (GT *ACes*) system involves the use of engineered target vectors carrying the therapeutic gene of interest. These are introduced onto the GT platform *ACes* via site-specific recombination. As with the CPP *ACes*, the use of engineered target vectors maximizes the use of the *de novo* generated GT platform
10 *ACes* for most gene targets. Furthermore, using lambda integrase technology, GT platform *ACes* containing multiple *attP* sites permits the opportunity to incorporate multiple therapeutic targets onto a single platform. This could be of value in cases where a defined therapy requires multiple gene targets, a single therapeutic target requires an
15 additional gene regulatory factor or a GT *ACes* requires a "kill" switch.

Similar to the CPP *ACes*, a feature of the gene target vector is the *promoterless* marker gene containing an upstream *attB* site (marker 2 on Figure 9). Normally, the marker (in this case, a cell surface antigen that can be sorted by FACS would be ideal) would not be expressed unless it
20 is placed downstream of a promoter sequence. Using the lambda integrase technology (λ INT_{E174R} on figure 9), site-specific recombination between the *attB* site on the vector and the promoter- *attP* site on the GT platform *ACes* results in the expression of marker#2 on the gene target vector, i.e. positive selection for the site-specific event. Site-specific
25 recombination events on the GT *ACes* versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR.

For expression of the therapeutic gene, human specific promoters, such as a ferritin heavy chain promoter (SEQ ID NO:128); EF1 α or RNA

-75-

Poll, are used. These promoters are for high level expression of a cDNA encoded therapeutic protein. In addition to expressing cDNA (or even hybrid cDNA/artificial intron constructs), the GT platform ACes are used for engineering and expressing large genomic fragments carrying

5 therapeutic genes of interest expressed from native promoter sequences. This is of importance in situations where the therapy requires precise cell specific expression or in instances where expression is best achieved from genomic clones versus cDNA.

10 **3. Selectable markers for use, for example, in Gene Therapy (GT)**

The following are selectable markers that can be incorporated into human ACes and used for selection.

15 **Dual Resistance to 4-Hydroperoxycyclophosphamide and Methotrexate by Retroviral Transfer of the Human Aldehyde Dehydrogenase Class 1 Gene and a Mutated Dihydrofolate Reductase Gene**

The genetic transfer of drug resistance to hematopoietic cells is one approach to overcoming myelosuppression caused by high-dose chemotherapy. Because cyclophosphamide (CTX) and methotrexate
20 (MTX) are commonly used non-cross-resistant drugs, generation of dual drug resistance in hematopoietic cells that allows dose intensification may increase anti-tumor effects and circumvent the emergence of drug-resistant tumors, a retroviral vector containing a human cytosolic ALDH-1-encoding DNA clone and a human doubly mutated DHFR-encoding
25 clone (Phe22/Ser31; termed F/S in the description of constructs) to generate increased resistance to CTX and MTX were constructed (Takebe *et al.* (2001) *Mol Ther* 3(1):88-96). This construct may be useful for protecting patients from high-dose CTX- and MTX-induced myelosuppression. ACes can be similarly constructed.

-76-

Multiple mechanisms of N-phosphonacetyl-L-aspartate resistance in human cell lines: carbamyl-P synthetase/aspartate transcarbamylase/dihydro-orotase gene amplification is frequent only when chromosome 2 is rearranged

5

Rodent cells resistant to N-phosphonacetyl-L-aspartate (PALA) invariably contain amplified carbamyl-P synthetase/aspartate transcarbamylase/dihydro-orotase (CAD) genes, usually in widely spaced tandem arrays present as extensions of the same chromosome arm that carries a single copy of CAD in normal cells (Smith *et al.* (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:1816-21). In contrast, amplification of CAD is very infrequent in several human tumor cell lines. Cell lines with minimal chromosomal rearrangement and with unrearranged copies of chromosome 2 rarely develop intrachromosomal amplifications of CAD.

10

15 These cells frequently become resistant to PALA through a mechanism that increases the aspartate transcarbamylase activity with no increase in CAD copy number, or they obtain one extra copy of CAD by forming an isochromosome 2p or by retaining an extra copy of chromosome 2. In cells with multiple chromosomal aberrations and rearranged copies of

20 chromosome 2, amplification of CAD as tandem arrays from rearranged chromosomes is the most frequent mechanism of PALA resistance. All of these different mechanisms of PALA resistance are blocked in normal human fibroblasts. Thus, ACes with multiple copies of the CAD gene would provide PALA resistance.

25

Retroviral coexpression of thymidylate synthase and dihydrofolate reductase confers fluoropyrimidine and antifolate resistance

Retroviral gene transfer of dominant selectable markers into hematopoietic cells can be used to select genetically modified cells in vivo or to attenuate the toxic effects of chemotherapeutic agents. Fantz *et al.* ((1998) *Biochem Biophys Res Comm* 243(1):6-12) have shown that

30

-77-

retroviral gene transfer of thymidylate synthase (TS) confers resistance to TS directed anticancer agents and that co-expression of TS and dihydrofolate reductase (DHFR) confers resistance to TS and DHFR cytotoxic agents. Retroviral vectors encoding *Escherichia coli* TS, human TS, and the Tyr-to-His at residue 33 variant of human TS (Y33HhTS) were constructed and fibroblasts transfected with these vectors conferred comparable resistance to the TS-directed agent fluorodeoxyuridine (FdUrd, approximately 4-fold). Retroviral vectors that encode dual expression of Y33HhTS and the human L22Y DHFR (L22YhDHFR) variants conferred resistance to FdUrd (3- to 5-fold) and trimetrexate (30- to 140-fold). A L22YhDHFR-Y33HhTS chimeric retroviral vector was also constructed and transduced cells were resistant to FdUrd (3-fold), AG337 (3-fold), trimetrexate (100-fold) and methotrexate (5-fold). These results show that recombinant retroviruses can be used to transfer the cDNA that encodes TS and DHFR and dual expression in transduced cells is sufficiently high to confer resistance to TS and DHFR directed anticancer agents. ACes can be similarly constructed.

Human CD34+ cells do not express glutathione S-transferases alpha

The expression of glutathione S-transferases alpha (GST alpha) in human hematopoietic CD34+ cells and bone marrow was studied using RT-PCR and immunoblotting (Czerwinski M, Kiem *et al.* (1997) *Gene Ther* 4(3):268-70). The GSTA1 protein conjugates glutathione to the stem cell selective alkylator busulfan. This reaction is the major pathway of elimination of the compound from the human body. Human hematopoietic CD34+ cells and bone marrow do not express GSTA1 message, which was present at a high level in liver, an organ relatively resistant to busulfan toxicity in comparison to bone marrow. Similarly, baboon CD34+ cells and dog bone marrow do not express GSTA1. Thus, human

-78-

GSTA1 is a chemoprotective selectable marker in human stem cell gene therapy and could be employed in ACes construction.

Selection of retrovirally transduced hematopoietic cells using CD24 as a marker of gene transfer

5 Pawliuk *et al.* ((1994) *Blood* 84(9):2868-2877) have investigated the use of a cell surface antigen as a dominant selectable marker to facilitate the detection and selection of retrovirally infected target cells. The small coding region of the human cell surface antigen CD24 (approximately 240 bp) was introduced into a myeloproliferative sarcoma virus (MPSV)-based retroviral vector, which was then used to infect day 4
10 5-fluorouracil (5-FU)-treated murine bone marrow cells. Within 48 hours of termination of the infection procedure CD24-expressing cells were selected by fluorescent-activated cell sorting (FACS) with an antibody directed against the CD24 antigen. Functional analysis of these cells
15 showed that they included not only *in vitro* clonogenic progenitors and day 12 colony-forming unit-spleen but also cells capable of competitive long-term hematopoietic repopulation. Double-antibody labeling studies performed on recipients of retrovirally transduced marrow cells showed that some granulocytes, macrophages, erythrocytes, and, to a lesser
20 extent, B and T lymphocytes still expressed the transduced CD24 gene at high levels 4 months later. No gross abnormalities in hematopoiesis were detected in mice repopulated with CD24-expressing cells. These results show that the use of the CD24 cell surface antigen as a retrovirally encoded marker permits rapid, efficient, and nontoxic selection *in vitro* of
25 infected primary cells, facilitates tracking and phenotyping of their progeny, and provides a tool to identify elements that regulate the expression of transduced genes in the most primitive hematopoietic cells. ACes could be similarly constructed.

-79-

DeltahGHR, a biosafe cell surface-labeling molecule for analysis and selection of genetically transduced human cells

A selectable marker for retroviral transduction and selection of human and murine cells is known (see, Garcia-Ortiz *et al.* (2000) *Hum Gene Ther* 11(2):333-46). The molecule expressed on the cell surface of the transduced population is a truncated version of human growth hormone receptor (deltahGHR), capable of ligand (hGH) binding, but devoid of the domains involved in signal triggering. The engineered molecule is stably expressed in the target cells as an inert protein unable to trigger proliferation or to rescue the cells from apoptosis after ligand binding. This new marker, has a wide application spectrum, since hGHR in the human adult is highly expressed only in liver cells, and lower levels have been reported in certain lymphocyte cell populations. The deltagHHR label has high biosafety potential, as it belongs to a well-characterized hormonal system that is nonessential in adults, and there is extensive clinical experience with hGH administration in humans. The differential binding properties of several monoclonal antibodies (MAbs) are used in a cell rescue method in which the antibody used to select deltagHHR-transduced cells is eluted by competition with hGH or, alternatively biotinylated hGH is used to capture tagged cells. In the latter system, the final purified population is recovered free of attached antibodies in hGH (a substance approved for human use)-containing medium. Such a system could be used to identify ACes containing cells.

4. Transgenic models for evaluation of genes and discovery of new traits in plants

Of interest is the use of plants and plant cells containing artificial chromosomes for the evaluation of new genetic combinations and discovery of new traits. Artificial chromosomes, by virtue of the fact that they can contain significant amounts of DNA can also therefore encode

-80-

numerous genes and accordingly a multiplicity of traits. It is contemplated here that artificial chromosomes, when formed from one plant species, can be evaluated in a second plant species. The resultant phenotypic changes observed, for example, can indicate the nature of the genes contained within the DNA contained within the artificial chromosome, and hence permit the identification of novel genetic activities. Artificial chromosomes containing euchromatic DNA or partially containing euchromatic DNA can serve as a valuable source of new traits when transferred to an alien plant cell environment. For example, it is contemplated that artificial chromosomes derived from dicot plant species can be introduced into monocot plant species by transferring a dicot artificial chromosome. The dicot artificial chromosome possessing a region of euchromatic DNA containing expressed genes.

The artificial chromosomes can be designed to allow the artificial chromosome to recombine with the naturally occurring plant DNA in such a fashion that a large region of naturally occurring plant DNA becomes incorporated into the artificial chromosome. This allows the artificial chromosome to contain new genetic activities and hence carry novel traits. For example, an artificial chromosome can be introduced into a wild relative of a crop plant under conditions whereby a portion of the DNA present in the chromosomes of the wild relative is transferred to the artificial chromosome. After isolation of the artificial chromosome, this naturally occurring region of DNA from the wild relative, now located on the artificial chromosome can be introduced into the domesticated crop species and the genes encoded within the transferred DNA expressed and evaluated for utility. New traits and gene systems can be discovered in this fashion. The artificial chromosome can be modified to contain sequences that promote homologous recombination within plant cells, or

-81-

be modified to contain a genetic system that functions as a site-specific recombination system.

Artificial chromosomes modified to recombine with plant DNA offer many advantages for the discovery and evaluation of traits in different
5 plant species. When the artificial chromosome containing DNA from one plant species is introduced into a new plant species, new traits and genes can be introduced. This use of an artificial chromosome allows for the ability to overcome the sexual barrier that prevents transfer of genes from one plant species to another species. Using artificial chromosomes in this
10 fashion allows for many potentially valuable traits to be identified including traits that are typically found in wild species. Other valuable applications for artificial chromosomes include the ability to transfer large regions of DNA from one plant species to another, such as DNA encoding potentially valuable traits such as altered oil, carbohydrate or protein
15 composition, multiple genes encoding enzymes capable of producing valuable plant secondary metabolites, genetic systems encoding valuable agronomic traits such as disease and insect resistance, genes encoding functions that allow association with soil bacterium such as growth promoting bacteria or nitrogen fixing bacteria, or genes encoding traits
20 that confer freezing, drought or other stress tolerances. In this fashion, artificial chromosomes can be used to discover regions of plant DNA that encode valuable traits.

The artificial chromosome can also be designed to allow the transfer and subsequent incorporation of these valuable traits now located
25 on the artificial chromosome into the natural chromosomes of a plant species. In this fashion the artificial chromosomes can be used to transfer large regions of DNA encoding traits normally found in one plant species into another plant species. In this fashion, it is possible to derive a plant cell that no longer needs to carry an artificial chromosome to

-82-

posses the novel trait. Thus, the artificial chromosome would serve as the transfer mechanism to permit the formation of plants with greater degree of genetic diversity.

The design of an artificial chromosome to accomplish the afore-
5 mentioned purposes can include within the artificial chromosome the presence of specific DNA sequences capable of acting as sites for homologous recombination to take place. For example, the DNA sequence of *Arabidopsis* is now known. To construct an artificial chromosome capable of recombining with a specific region of *Arabidopsis*
10 DNA, a sequence of *Arabidopsis* DNA, normally located near a chromosomal location encoding genes of potential interest can be introduced into an artificial chromosome by methods provided herein. It may be desirable to include a second region of DNA within the artificial chromosome that provides a second flanking sequence to the region
15 encoding genes of potential interest, to promote a double recombination event which would ensure transfer of the entire chromosomal region, encoding genes of potential interest, to the artificial chromosome. The modified artificial chromosome, containing the DNA sequences capable of homologous recombination region, can then be introduced into
20 *Arabidopsis* cells and the homologous recombination event selected.

It is convenient to include a marker gene to allow for the selection of a homologous recombination event. The marker gene is preferably inactive unless activated by an appropriate homologous recombination event. For example, US 5,272,071, describes a method where an
25 inactive plant gene is activated by a recombination event such that desired homologous recombination events can be easily scored. Similarly, US 5,501,967 describes a method for the selection of homologous recombination events by activation of a silent selection gene first introduced into the plant DNA, the gene being activated by an appropriate

-83-

homologous recombination event. Both of these methods can be applied to enable a selective process to be included to select for recombination between an artificial chromosome and a plant chromosome. Once the homologous recombination event is detected, the artificial chromosome, once selected, is isolated and introduced into a recipient cell, for example, tobacco, corn, wheat or rice, and the expression of the newly introduced DNA sequences evaluated.

Phenotypic changes in the recipient plant cells containing the artificial chromosome, or in regenerated plants containing the artificial chromosome, allows for the evaluation of the nature of the traits encoded by the *Arabidopsis* DNA, under conditions naturally found in plant cells, including the naturally occurring arrangement of DNA sequences responsible for the developmental control of the traits in the normal chromosomal environment.

Traits such as durable fungal or bacterial disease resistance, new oil and carbohydrate compositions, valuable secondary metabolites such as phytosterols, flavonoids, efficient nitrogen fixation or mineral utilization, resistance to extremes of drought, heat or cold are all found within different populations of plant species and are often governed by multiple genes. The use of single gene transformation technologies does not permit the evaluation of the multiplicity of genes controlling many valuable traits. Thus, incorporation of these genes into artificial chromosomes allows the rapid evaluation of the utility of these genetic combinations in heterologous plant species.

The large scale order and structure of the artificial chromosome provides a number of unique advantages in screening for new utilities or novel phenotypes within heterologous plant species. The size of new DNA that can be carried by an artificial chromosome can be millions of base pairs of DNA, representing potentially numerous genes that may

-84-

have novel utility in a heterologous plant cell. The artificial chromosome is a "natural" environment for gene expression, the problems of variable gene expression and silencing seen for genes transferred by random insertion into a genome should not be observed. Similarly, there is no
5 need to engineer the genes for expression, and the genes inserted would not need to be recombinant genes. Thus, one expects the expression from the transferred genes to be temporal and spatial, as observed in the species from where the genes were initially isolated. A valuable feature for these utilities is the ability to isolate the artificial chromosomes and to
10 further isolate, manipulate and introduce into other cells artificial chromosomes carrying unique genetic compositions.

Thus, the use of artificial chromosomes and homologous recombination in plant cells can be used to isolate and identify many valuable crop traits.

15 In addition to the use of artificial chromosomes for the isolation and testing of large regions of naturally occurring DNA, methods for the use of artificial chromosomes and cloned DNA are also contemplated. Similar to that described above, artificial chromosomes can be used to carry large regions of cloned DNA, including that derived from other plant species.

20 The ability to incorporate novel DNA elements into an artificial chromosome as it is being formed allows for the development of artificial chromosomes specifically engineered as a platform for testing of new genetic combinations, or "genomic" discoveries for model species such as *Arabidopsis*. It is known that specific "recombinase" systems can be
25 used in plant cells to excise or re-arrange genes. These same systems can be used to derive new gene combinations contained on an artificial chromosome.

The artificial chromosomes can be engineered as platforms to accept large regions of cloned DNA, such as that contained in Bacterial

-85-

Artificial Chromosomes (BACs) or Yeast Artificial Chromosomes (YACs). It is further contemplated, that as a result of the typical structure of artificial chromosomes containing tandemly repeated DNA blocks, that sequences other than cloned DNA sequence can be introduced by recombination processes. In particular recombination within a predefined region of the tandemly repeated DNA within the artificial chromosome provides a mechanism to "stack" numerous regions of cloned DNA, including large regions of DNA contained within BACs or YACs clones. Thus, multiple combinations of genes can be introduced onto artificial chromosomes and these combinations tested for functionality. In particular, it is contemplated that multiple YACs or BACs can be stacked onto an artificial chromosomes, the BACs or YACs containing multiple genes of complex pathways or multiple genetic pathways. The BACs or YACs are typically selected based on genetic information available within the public domain, for example from the Arabidopsis Information Management System (<http://aims.cps.msu.edu/aims/index.html>) or the information related to the plant DNA sequences available from the Institute for Genomic Research (<http://www.tigr.org>) and other sites known to those skilled in the art. Alternatively, clones can be chosen at random and evaluated for functionality. It is contemplated that combinations providing a desired phenotype can be identified by isolation of the artificial chromosome containing the combination and analyzing the nature of the inserted cloned DNA.

In this regard, it is contemplated that the use of site-specific recombination sequences can have considerable utility in developing artificial chromosomes containing DNA sequences recognized by recombinase enzymes and capable of accepting DNA sequences containing same. The use of site-specific recombination as a means to target an introduced DNA to a specific locus has been demonstrated in

-86-

the art and such methods can be employed. The recombinase systems can also be used to transfer the cloned DNA regions contained within the artificial chromosome to the naturally occurring plant or mammalian chromosomes.

- 5 As noted herein, many site-specific recombinases are known and can be identified (Kilby *et al.* (1993) *Trends in Genetics* 9:413-418). The three recombinase systems that have been extensively employed include: an activity identified as R encoded by the pSR1 plasmid of *Zygosaccharomyces rouxii*, FLP encoded for the 2 μ m circular plasmid from
10 *Saccharomyces cerevisiae* and *Cre-lox* from the phage P1.

- The integration function of site-specific recombinases is contemplated as a means to assist in the derivation of genetic combinations on artificial chromosomes. In order to accomplish this, it is contemplated that a first step of introducing site-specific recombinase
15 sites into the genome of a plant cell in an essentially random manner is conducted, such that the plant cell has one or more site-specific recombinase recognition sequences on one or more of the plant chromosomes. An artificial chromosome is then introduced into the plant cell, the artificial chromosome engineered to contain a recombinase
20 recognition site (e.g., integration site) capable of being recognized by a site-specific recombinase. Optionally, a gene encoding a recombinase enzyme is also included, preferably under the control of an inducible promoter. Expression of the site-specific recombinase enzyme in the plant cell, either by induction of a inducible recombinase gene, or
25 transient expression of a recombinase sequence, causes a site-specific recombination event to take place, leading to the insertion of a region of the plant chromosomal DNA (containing the recombinase recognition site) into the recombinase recognition site of the artificial chromosome, and forming an artificial chromosome containing plant chromosomal DNA.

-87-

The artificial chromosome can be isolated and introduced into a heterologous host, preferably a plant host, and expression of the newly introduced plant chromosomal DNA can be monitored and evaluated for desirable phenotypic changes. Accordingly, carrying out this

5 recombination with a population of plant cells wherein the chromosomally located recombinase recognition site is randomly scattered throughout the chromosomes of the plant, can lead to the formation of a population of artificial chromosomes, each with a different region of plant chromosomal DNA, and each potentially representing a novel genetic combination.

10 This method requires the precise site-specific insertion of chromosomal DNA into the artificial chromosome. This precision has been demonstrated in the art. For example, Fukushige and Sauer ((1992) Proc. Natl. Acad. Sci. USA, 89:7905-7909) demonstrated that the *Cre-lox* homologous recombination system could be successfully employed to

15 introduce DNA into a predefined locus in a chromosome of mammalian cells. In this demonstration a promoter-less antibiotic resistance gene modified to include a *lox* sequence at the 5' end of the coding region was introduced into CHO cells. Cells were re-transformed by electroporation with a plasmid that contained a promoter with a *lox* sequence and a

20 transiently expressed *Cre* recombinase gene. Under the conditions employed, the expression of the *Cre* enzyme catalyzed the homologous recombination between the *lox* site in the chromosomally located promoter-less antibiotic resistance gene, and the *lox* site in the introduced promoter sequence, leading to the formation of a functional antibiotic

25 resistance gene. The authors demonstrated efficient and correct targeting of the introduced sequence, 54 of 56 lines analyzed corresponded to the predicted single copy insertion of the DNA due to *Cre* catalyzed site-specific homologous recombination between the *lox* sequences.

-88-

Accordingly a *lox* sequence may be first added to a genome of a plant species capable of being transformed and regenerated to a whole plant to serve as a recombinase target DNA sequence for recombination with an artificial chromosome. The *lox* sequence may be optimally
5 modified to further contain a selectable marker which is inactive but can be activated by insertion of the *lox* recombinase recognition sequence into the artificial chromosome.

A promoterless marker gene or selectable marker gene linked to the recombinase recognition sequence, which is first inserted into the
10 chromosomes of a plant cell can be used to engineer a platform chromosome. A promoter is linked to a recombinase recognition site, in an orientation that allows the promoter to control the expression of the marker or selectable marker gene upon recombination within the artificial chromosome. Upon a site-specific recombination event between a
15 recombinase recognition site in a plant chromosome and the recombinase recognition site within the introduced artificial chromosome, a cell is derived with a recombined artificial chromosome, the artificial chromosome containing an active marker or selectable marker activity that permits the identification and or selection of the cell.

20 The artificial chromosomes can be transferred to other plant or animal species and the functionality of the new combinations tested. The ability to conduct such an inter-chromosomal transfer of sequences has been demonstrated in the art. For example, the use of the *Cre-lox* recombinase system to cause a chromosome recombination event
25 between two chromatids of different chromosomes has been shown.

Any number of recombination systems may be employed as described herein, such as, but not limited to, bacterially derived systems such as the att/int system of phage lambda, and the Gin/gix system.

-89-

More than one recombination system may be employed, including, for example, one recombinase system for the introduction of DNA into an artificial chromosome, and a second recombinase system for the subsequent transfer of the newly introduced DNA contained within an
5 artificial chromosome into the naturally occurring chromosome of a second plant species. The choice of the specific recombination system used will be dependent on the nature of the modification contemplated.

By having the ability to isolate an artificial chromosome, in particular, artificial chromosomes containing plant chromosomal DNA
10 introduced via site-specific recombination, and re-introduce the chromosome into other mammalian or plant cells, particularly plant cells, these new combinations can be evaluated in different crop species without the need to first isolate and modify the genes, or carry out multiple transformations or gene transfers to achieve the same
15 combination isolation and testing combinations of the genes in plants. The use of a site-specific recombinase also allows the convenient recovery of the plant chromosomal region into other recombinant DNA vectors and systems, such as mammalian or insect systems, for manipulation and study.

20 Also contemplated herein are *ACes*, cell lines and methods for use in screening a new chromosomal combinations, deletions, truncations with eucaryotic genome that take advantage of the site-specific recombination systems incorporated onto platform *ACes* provided herein. For example, provided herein is a cell line useful for making a library of
25 *ACes*, comprising a multiplicity of heterologous recombination sites randomly integrated throughout the endogenous chromosomes. Also provided herein is a method of making a library of *ACes* comprising random portions of a genome, comprising introducing one or more *ACes* into a cell line comprising a multiplicity of heterologous recombination

-90-

sites randomly integrated throughout the endogenous chromosomes, under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites within the cell's chromosomal DNA; and isolating said

5 multiplicity of *ACes*, thereby producing a library of *ACes* whereby multiple *ACes* have different portions of the genome within. Also provided herein is a library of cells useful for genomic screening, said library comprising a multiplicity of cells, wherein each cell comprises an *ACes* having a mutually exclusive portion of a chromosomal nucleic acid therein. The

10 library of cells can be from a different species and/or cell type than the chromosomal nucleic acid within the *ACes*. Also provided is a method of making one or more cell lines, comprising

- a) integrating into endogenous chromosomal DNA of a selected cell species, a multiplicity of heterologous recombination sites,
- 15 b) introducing a multiplicity of *ACes* under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites integrated within the cell's endogenous chromosomal DNA;
- c) isolating said multiplicity of *ACes*, thereby producing a library of
- 20 *ACes* whereby a multiplicity of *ACes* have mutually exclusive portions of the endogenous chromosomal DNA therein;
- d) introducing the isolated multiplicity of *ACes* of step c) into a multiplicity of cells, thereby creating a library of cells;
- e) selecting different cells having mutually exclusive *ACes* therein
- 25 and clonally expanding or differentiating said different cells into clonal cell cultures, thereby creating one or more cell lines.

These *ACes*, cell lines and methods utilize the site-specific recombination sites on platform *ACes* analogous YAC manipulation related to: the methods of generating terminal deletions in normal and

-91-

artificial chromosomes (e.g., *ACes*; as described in Vollrath et al., 1988, *PNAS, USA*, 85:6027-66031; and Pavan et al., *PNAS, USA*, 87:1300-1304); the methods of generating interstitial deletions in normal and artificial chromosomes (as described in Campbell et al., 1991, *PNAS, USA*, 888:5744-5748); and the methods of detecting homologous recombination between two *ACes* (as described in Cellini et al., 1991, *Nuc. Acid Res.*, 19(5):997-1000).

5. Use of platform *ACes* in Pharmacogenomic/toxicology applications (development of "Reporter *ACes*")

10 In addition to the placement of genes onto *ACes* chromosomes for therapeutic protein production or gene therapy, the platform can be engineered via the *IntR* lambda integrase to carry reporter-linked constructs (reporter genes) that monitor changes in cellular physiology as measured by the particular reporter gene (or a series of different reporter
15 genes) readout. The reporter linked constructs are designed to include a gene that can be detected (by for example fluorescence, drug resistance, immunohistochemistry, or transcript production, and the like) with well-known regulatory sequences that would control the expression of the detectable gene. Exemplary regulatory promoter sequences are well-
20 known in the art:

A) Reporter *ACes* for drug pathway screening

The *ACes* can be engineered to carry reporter-linked constructs that indicate a signal is being transduced through one or a number of pathways. For example, transcriptionally regulated promoters from genes
25 at the end (or any other chosen point) of particular signal transduction pathways could be engineered on the *ACes* to express the appropriate readout (either by fluorescent protein production or drug resistance) when the pathway is activated (or down-regulated as well). In one embodiment, a number of reporters from different can be placed on a

-92-

ACes chromosome. Cells (and/or whole animals) containing such a Reporter *ACes* could be exposed to a variety of drugs or compounds and monitored for the effects of the drugs or compounds upon the selected pathway(s) by the reporter gene(s). Thus, drugs or compounds can be classified or identified by particular pathways they excite or down-regulate. Similarly, transcriptional profiles obtained from genomic array experiments can be biologically validated using the reporter *ACes* provided herein.

B) Reporter *ACes* for toxic compound testing

Environmental or man-made genotoxicants can be tested in cell lines carrying a number of reporter-genes platform *ACes* linked to promoters that are transcriptionally regulated in response to DNA damage, induced apoptosis or necrosis, and cell-cycle perturbations. Furthermore, new drugs and/or compounds could be tested in a similar manner with the genotoxicant *ACes* reporter for their cellular/genetic toxicity by such a screen. Likewise, toxic compound testing could be carried out in whole transgenic animals carrying the *ACes* chromosome that measures genotoxicant exposure ("canary in a coal mine"). Thus, the same or similar type *ACes* could be used for toxicity testing in either a cell-based or whole animal setting. An example would include *ACes* that carry reporter-linked genes controlled by various cytochrome P450 profiled promoters and the like.

C) Reporter *ACes* for individualized pharmacogenomics/drug profiling

A common disease may arise via various mechanisms. In many instances there are multiple treatments available for a given disease. However, the success of a given treatment may depend upon the mechanism by which the disease originated and/or by the genetic background of the patient. In order to establish the most effective

-93-

treatment for a given patient one could utilize the *ACes* reporters provided herein. *ACes* reporters can be used in patient cell samples to determine an individualized drug regimen for the patient. In addition, potential polymorphisms affecting the transcriptional regulation of an individual's particular gene can be assessed by this approach.

D) Reporter *ACes* for classification of similar patient tumors

As with other diseases as described in 5.C) above, cancer cells arise via different mechanisms. Furthermore, as a cancerous cell propagates it may undergo genomic alterations. An *ACes* reporter transferred to cells of different patients having the same disease, i.e. similar cancers, could be used to categorize the particular cancer of each patient, thereby facilitating the identification of the most effective therapeutic regimen. Examples would include the validation of array profiling of certain classes of breast cancers. Subsequently, appropriate drug profiling could be carried out as described above.

E) Reporter *ACes* as a "differentiation" sensor

Using the *ACes* reporter as a "differentiation" sensor in stem cells or other progenitor cells in order to enrich by selection (either FACS based screening, drug selection and/or use of suicide gene) for a particular class of differentiated or undifferentiated cells. For example, in one embodiment, this assay could also be used for compound screening for small molecule modifiers of cell differentiation.

F) Whole animal studies with Reporter *ACes*

Finally, with whole-body fluorescence imaging technology (Yang et al. (2000) PNAS 97:12278) any of the above Reporter *ACes* methods could be used in conjunction with whole-body imaging to monitor reporter genes within whole animals without sacrificing the animal. This would allow temporal and spatial analysis of expression patterns under a given set of conditions. The conditions tested may include for example, normal

-94-

differentiation of a stem cell, response to drug or compound treatment whether targeted to the diseased tissue or presented systemically, response to genotoxics, and the like.

The following examples are included for illustrative purposes only
5 and are not intended to limit the scope of the invention.

EXAMPLE 1

pFK161

Cosmid pFK161 (SEQ ID NO: 118) was obtained from Dr. Gyula Hadlaczky and contains a 9 kb *NotI* insert derived from a murine rDNA
10 repeat (see clone 161 described in PCT Application Publication No. WO97/40183 by Hadlaczky *et al.* for a description of this cosmid). This cosmid, referred to as clone 161 contains sequence corresponding to nucleotides 10,232-15,000 in SEQ ID NO. 26. It was produced by inserting fragments of the megachromosome (see, U.S. Patent No.
15 6,077,697 and International PCT application No. WO 97/40183). For example, H1D3, which was deposited at the European Collection of Animal Cell Culture (ECACC) under Accession No. 96040929, is a mouse-hamster hybrid cell line carrying this megachromosome into plasmid pWE15 (Stratagene, La Jolla, California; SEQ ID No. 31) as
20 follows. Half of a 100 μ l low melting point agarose block (mega-plug) containing isolated SATACs was digested with *NotI* overnight at 37°C. Plasmid pWE15 was similarly digested with *NotI* overnight. The mega-plug was then melted and mixed with the digested plasmid, ligation buffer and T4 DNA ligase. Ligation was conducted at 16°C overnight. Bacterial
25 DH5 α cells were transformed with the ligation product and transformed cells were plated onto LB/Amp plates. Fifteen to twenty colonies were grown on each plate for a total of 189 colonies. Plasmid DNA was isolated from colonies that survived growth on LB/Amp medium and analyzed by Southern blot hybridization for the presence of DNA that

-95-

hybridized to a pUC19 probe. This screening methodology assured that all clones, even clones lacking an insert but yet containing the pWE15 plasmid, would be detected.

Liquid cultures of all 189 transformants were used to generate
5 cosmid minipreps for analysis of restriction sites within the insert DNA. Six of the original 189 cosmid clones contained an insert. These clones were designated as follows: 28 (~9-kb insert), 30 (~9-kb insert), 60 (~4-kb insert), 113 (~9-kb insert), 157 (~9-kb insert) and 161 (~9-kb insert). Restriction enzyme analysis indicated that three of the clones
10 (113, 157 and 161) contained the same insert.

For sequence analysis the insert of cosmid clone no. 161 was subcloned as follows. To obtain the end fragments of the insert of clone no. 161, the clone was digested with *NotI* and *BamHI* and ligated with *NotI/BamHI*-digested pBluescript KS (Stratagene, La Jolla, California).
15 Two fragments of the insert of clone no. 161 were obtained: a 0.2-kb and a 0.7-kb insert fragment. To subclone the internal fragment of the insert of clone no. 161, the same digest was ligated with *BamHI*-digested pUC19. Three fragments of the insert of clone no. 161 were obtained: a 0.6-kb, a 1.8-kb and a 4.8-kb insert fragment.

20 The insert corresponds to an internal section of the mouse ribosomal RNA gene (rDNA) repeat unit between positions 7551-15670 as set forth in GENBANK accession no. X82564, which is provided as SEQ ID NO. 18. The sequence data obtained for the insert of clone no. 161 is set forth in SEQ ID NOS. 19-25. Specifically, the individual
25 subclones corresponded to the following positions in GENBANK accession no. X82564 (SEQ ID NO:18) and in SEQ ID NOS. 19-25:

-96-

Subclone	Start	End	Site	SEQ ID No.
	in X82564			
161k1	7579	7755	<i>NotI, BamHI</i>	19
161m5	7756	8494	<i>BamHI</i>	20
5 161m7	8495	10231	<i>BamHI</i>	21 (shows only sequence corresponding to nt. 8495-8950), 22 (shows only sequence corresponding to nt. 9851- 10231)
161m12	10232	15000	<i>BamHI</i>	23 (shows only sequence corresponding to nt. 10232-10600), 24 (shows only sequence corresponding to nt. 14267-15000)
161k2	15001	15676	<i>NotI, BamHI</i>	25

The sequence set forth in SEQ ID NOs. 19-25 diverges in some
 10 positions from the sequence presented in positions 7551-15670 of
 GENBANK accession no. X82564. Such divergence may be attributable
 to random mutations between repeat units of rDNA.

For use herein, the rDNA insert from the clone was prepared by
 digesting the cosmid with *NotI* and *BglII* and was purified as described
 15 above. Growth and maintenance of bacterial stocks and purification of
 plasmids were performed using standard well known methods (see, *e.g.*,
 Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, 2nd
 Edition, Cold Spring Harbor Laboratory Press), and plasmids were purified
 from bacterial cultures using Midi - and Maxi-preps Kits (Qiagen,
 20 Mississauga, Ontario).

pDsRed1N1

This vector is available from Clontech (see SEQ ID No. 29) and
 encodes the red fluorescent protein (DsRed; Genbank accession no.
 AF272711; SEQ ID Nos. 39 and 40). DsRed, which has a vivid red
 25 fluorescence, was isolated from the IndoPacific sea anemone relative
Discosoma species. The plasmid pDsRed1N1 (Clontech; SEQ ID No. 29)
 constitutively expresses a human codon-optimized variant of the

-97-

fluorescent protein under control of the CMV promoter. Unmodified, this vector expresses high levels of DsRed1 and includes sites for creating N-terminal fusions by cloning proteins of interest into the multiple cloning site (MCS). It is Kan and Neo resistant for selection in bacterial or
5 eukaryotic cells.

Plasmid pMG

Plasmid pMG (InvivoGen, San Diego, California; see SEQ. ID. NO. 27 for the nucleotide sequence of pMG) contains the hygromycin phosphotransferase gene under the control of the immediate-early human
10 cytomegalovirus (hCMV) enhancer/promoter with intron A. Vector pMG also contains two transcriptional units allowing for the coexpression of two heterologous genes from a single vector sequence.

The first transcriptional unit of pMG contains a multiple cloning site for insertion of a gene of interest, the hygromycin phosphotransferase
15 gene (*hph*) and the immediate-early human cytomegalovirus (hCMV) enhancer/promoter with intron A (see, *e.g.*, Chapman *et al.* (1991) *Nuc. Acids Res.* 19:3979-3986) located upstream of *hph* and the multiple cloning site, which drives the expression of *hph* and any gene of interest inserted into the multiple cloning site as a polycistronic mRNA. The first
20 transcriptional unit also contains a modified EMCV internal ribosomal entry site (IRES) upstream of the *hph* gene but downstream of the hCMV promoter and MCS for ribosomal entry in translation of the *hph* gene (see SEQ ID NO. 27, nucleotides 2736-3308). The IRES is modified by insertion of the constitutive *E. coli* promoter (EM7) within an intron (IM7)
25 into the end of the IRES. In mammalian cells, the *E. coli* promoter is treated as an intron and is spliced out of the transcript. A polyadenylation signal from the bovine growth hormone (bGh) gene (see, *e.g.*, Goodwin and Rottman (1992) *J. Biol. Chem.* 267:16330-16334) and a pause site derived from the 3' flanking region of the human $\alpha 2$

-98-

globin gene (see, *e.g.*, Enriquez-Harris *et al.* (1991) *EMBO J.* 10:1833-1842) are located at the end of the first transcription unit. Efficient polyadenylation is facilitated by inserting the flanking sequence of the bGh gene 3' to the standard AAUAAA hexanucleotide sequence.

- 5 The second transcriptional unit of pMG contains another multiple cloning site for insertion of a gene of interest and an EF-1 α /HTLV hybrid promoter located upstream of this multiple cloning site, which drives the expression of any gene of interest inserted into the multiple cloning site. The hybrid promoter is a modified human elongation factor-1 alpha (EF-1
- 10 alpha) gene promoter (see, *e.g.*, Kim *et al.* (1990) *Gene* 91:217-223) that includes the R segment and part of the U5 sequence (R-U5') of the human T-cell leukemia virus (HTLV) type I long terminal repeat (see, *e.g.*, Takebe *et al.* (1988) *Mol. Cell. Biol.* 8:466-472). The Simian Virus 40 (SV40) late polyadenylation signal (see Carswell and Alwine (1989) *Mol.*
- 15 *Cell. Biol.* 9:4248-4258) is located downstream of the multiple cloning site. Vector pMG contains a synthetic polyadenylation site for the first and second transcriptional units at the end of the transcriptional unit based on the rabbit β -globin gene and containing the AATAAA hexanucleotide sequence and a GT/T-rich sequence with 22-23
- 20 nucleotides between them (see, *e.g.*, Levitt *et al.* (1989) *Genes Dev.* 3:1019-1025). A pause site derived from the C2 complement gene (see, Moreira *et al.* (1995) *EMBO J.* 14:3809-3819) is also located at the 3' end of the second transcriptional unit.

- Vector pMG also contains an ori sequence (ori pMB1) located
- 25 between the SV40 polyadenylation signal and the synthetic polyadenylation site.

EXAMPLE 2

A. Construction of targeting vector and transfection into LMtk- cells for the generation of platform chromosomes

-99-

- A targeting vector derived from the vector pWE15 (GeneBank Accession # X65279) was modified by replacing the *Sa*I (Klenow filled)/*Sma*I neomycin resistance containing fragment with the *Pvu*II/*Bam*HI (Klenow filled) puromycin resistance containing fragment
- 5 (isolated from plasmid pPUR, Clontech Laboratories, Inc. Palo Alto, CA; SEQ ID No. 30) resulting in plasmid pWEPuro. Subsequently a 9 Kb *Not*I fragment from the plasmid pFK161 (SEQ ID NO: 118) containing a portion of the mouse rDNA region was cloned into the *Not*I site of pWEPuro resulting in plasmid pWEPuro9K (Figure 2). The vector pWEPuro9K was
- 10 digested with *Spe*I to linearize and transfected into LMtk- mouse cells. Puromycin resistant colonies were isolated and subsequently tested for artificial chromosome formation via fluorescent *in situ* hybridization (FISH) (using mouse major and minor DNA repeat sequences, the puromycin gene and telomeres sequences as probes), and fluorescent activated cell
- 15 sorting (FACS). From this sort, a subclone was isolated containing an artificial chromosome, designated 5B11.12, which carries 4-8 copies of the puromycin resistance gene contained on the pWEPuro9K vector. FISH analysis of the 5B11.12 subclone demonstrated the presence of telomeres and mouse minor on the ACes. DOT PCR has been done on
- 20 the 5B11.12 ACes revealing the absence of uncharacterized euchromatic regions on the ACes. A recombination site, such as an *att* or *loxP* engineering site or a plurality thereof, was introduced onto this ACes thereby providing a platform for site-specific introduction of heterologous nucleic acid.
- 25 **B. Targeting a single sequence specific recombination site onto platform chromosomes**

After the generation of the 5B11.12 platform, a single sequence-specific recombination site is placed onto the platform chromosome via homologous recombination. For this, DNA sequences containing the site-

-100-

specific recombination sequence can be flanked with DNA sequences of homology to the platform chromosome. For example, using the platform chromosome made from the pWEPuro9K vector, mouse rDNA sequences or mouse major satellite DNA can be used as homologous sequences to target onto the platform chromosome. A vector is designed to have these homologous sequences flanking the site-specific recombination site and, after the appropriate restriction enzyme digest to generate free ends of homology to the platform chromosome, the DNA is transfected into cells harboring the platform chromosome (Figure 3). Examples of site-specific cassettes that are targeted to the platform chromosome using either mouse rDNA or mouse major repeat DNA include the SV40-attP-hygro cassette and a red fluorescent protein (RFP) gene flanked by loxP sites (Cre/lox, see, *e.g.*, U.S. Patent No. 4,959,317 and description herein). After transfection and integration of the site-specific cassette, homologous recombination events onto the platform chromosome are subcloned and identified by FACS (*e.g.* screen and single cell subclone via expression of resistance or fluorescent marker) and PCR analysis.

For example, a vector can be constructed containing regions of the mouse rDNA locus flanking a gene cassette containing the SV40 early reporter-bacteriophage lambda attP site-hygromycin selectable marker (see Figure 4 and described below). The use of the bacteriophage lambda attP site for lambda integrase-mediated site-specific recombination is described below. Homologous recombination event of the SV40-attP-hygro cassette onto the platform chromosome was identified using PCR primers that detect the homologous recombination and further confirmed by FISH analysis. After identifying subcloned colonies containing the platform chromosome with a single site-specific recombination site, cells carrying the platform chromosome with a single site-specific

-101-

recombination site can now be engineered with site-specific recombinases (e.g. lambda INT, Cre) for integrating a target gene expression vector.

**C. Targeting a red fluorescent protein (RFP) gene flanked by loxP sites
5 onto 5B11.12 platform**

As another example, while loxP recombination sites could have been introduced onto the *ACes* during *de novo* biosynthesis, it was thought that this might result in multiple segments of the *ACes* containing a high number of loxP sites, potentially leading to instability upon Cre-mediated recombination. A gene targeting approach was therefore
10 devised to introduce a more limited number of loxP recombination sites into a locus of the 5B11-12 *ACes* containing introduced and possibly co-amplified endogenous rDNA sequences. Although there are more than 200 copies of rDNA genes in the haploid mouse genome distributed
15 amongst 5-11 chromosomes (depending on strain), rDNA sequences were chosen as the target on the *ACes* since they represent a less frequent target than that of the satellite repeat sequences. Moreover, having observed much stronger pWEPuro9K hybridization to the 5B11-12 *ACes* than to other LMTK⁻ chromosomes and in light of the observation that the
20 transcribed spacer sequences within the rDNA may be less conserved than the rRNA coding regions, it was contemplated that a targeting vector based on the rDNA gene segment in pWEPuro9K would have a higher probability of targeting to the *ACes* rather than to other LMTK⁻ chromosomes. Accordingly, a targeting vector, pBSFKLoxDsRedLox, was
25 designed and constructed based on the rDNA sequences contained in pWEPuro9K.

The plasmid pBSFKLoxDsRedLox was generated in 4 steps. First, the *NotI* rDNA insert of pWEPuro9K (Figure 2) was inserted into pBS SK- (Stratagene) giving rise to pBSFK. Second, a loxP polylinker cassette was

-102-

generated by PCR amplification of pNEB193 (SEQ ID NO:32; New England Biolabs) using primers complementary to the M13 forward and reverse priming sites at their 3' end and a 34 bp 5' extension comprising a LoxP site. This cassette was reinserted into pNEB193 generating
5 p193LoxMCSLox. Third, the DsRed gene from pDsRed1-N1 (SEQ ID NO:29; Clontech) was then cloned into the polylinker between the loxP sites generating p193LoxDsRedLox. Fourth, a fragment consisting of the DsRed gene flanked by loxP sites was cloned into a unique *NdeI* within the rDNA insert of pBSFK generating pBSFKLoxDsRedLox.

10 A gel purified 11 Kb *PmlI* / *EcoRV* fragment of pBSFKLoxDsRedLox was used for transfection. To detect targeted integration, PCR primers were designed from rDNA sequences within the 5' *NotI*-*PmlI* fragment of pWEPuro9K that is not present on the targeting fragment (5' primer) and sequence within the LoxDsRedLox cassette (3' primer). If the targeting
15 DNA integrated correctly within the rDNA sequences, PCR amplification using these primers would give rise to a 2.3 Kb band. PCR reactions containing 1-4 μ l of genomic DNA were carried out according to the MasterTaq protocol (Eppendorf), using murine rDNA 5' primer (5'-CGGACAATGCGGTTGTGCGT-3'; SEQ ID NO:72) and DsRed 3' primer
20 (5'-GGCCCCGTAATGCAGAAGAA-3'; SEQ ID NO:73) and PCR products were analyzed by agarose gel electrophoresis.

1.5X10⁶ 5B11-12 LMTK⁻ cells were transfected with 2 μ g of the pBSFKLoxDsRedLox targeting DNA described above using Lipofectamine Plus (Invitrogen). For flow sorting, harvested cells were suspended in
25 medium and applied to the Becton Dickinson Vantage SE cell sorter, equipped with 488 nm lasers for excitation and 585/42 bandpass filter for optimum detection of RFP fluorescence. Cells were sorted using dPBS as sheath buffer. Negative control parental 5B11-12 cells and a positive control LMTK⁻ cell line stably transfected with DsRed were used to

-103-

establish the selection gates. The RFP positive gated populations were recovered, diluted in medium supplemented with 1X penicillin-streptomycin (Invitrogen), then plated and cultured as previously described. After 4 rounds of enrichment, the percentage of RFP positive

5 cells reached levels of 50% or higher. DNA from populations was analyzed by PCR for evidence of targeted integration. Ultimately, single cell subclones were established from positive pools and were analyzed by PCR and PCR-positive clones confirmed by FISH as described below. DNA was purified from pools or single cell clones using previously

10 described methods set forth in Lahm et al., Transgenic Res., 1998; 7:131-134, or in some cases using a Wizard Genomic DNA purification kit (Promega). For FISH analysis, a biotinylated DsRed gene probe was generated by PCR using DsRed specific primers and biotin-labeled dUTP (5' RFP primer: 5'-GGTTTAAAGTGCGCTCCTCCAAGAACGTCATC-3',

15 SEQ ID NO:74; and 3' RFP primer: 5'AGATCTAGAGCCGCGCTACAGGAACAGGTGGTGGCGGCC-3'; SEQ ID NO:75). To maximize the signal intensity of the DsRed probe, Tyramide amplification was carried out according to the manufacturers protocols (NEN).

20 The process of testing the feasibility of a more general targeting strategy that would not rely on enrichment *via* drug selection of stably transfected clones can be summarized as follows. A red fluorescent protein gene (RFP; encoded by the DsRed gene) was inserted between the loxP sites of the targeting vector to form pBSFKLoxDsRedLox. After

25 transfection with pBSFKLoxDsRedLox, sequential rounds of high speed flow sorting and expansion of sorted cells in culture could then be used to enrich for stable transformants expressing RFP. In the event of targeted integration, PCR screening with primers that amplify from a spacer region within the segment of the 45s pre-rRNA gene in pWEPuro9K to a specific

-104-

anchor sequence within the DsRed gene in the targeting cassette would give rise to a diagnostic 2.3 Kb band. However, as rDNA clusters are found on several chromosomes, confirmation of targeting to an *ACes* would require fluorescence in situ hybridization (FISH) analysis. Finally, the flanking of the DsRed gene by loxP sites would allow for its removal and subsequent replacement with other genes of interest.

After transfection of the targeting sequence into 5B11-12 cells, enrichment for targeted clones was carried out using a combination of flow cytometry to detect red-fluorescing cells and PCR screening.

10 Ultimately 17 single cell subclones were identified as potential targeted clones by PCR and of these 16 were found by FISH to contain the DsRed integration event into the *ACes*. These subclones are referred to herein as D11-C4, D11-C12, D11-H3, C9-C9, C9-B9, C9-F4, C9-H8, C9-F2, C9-G8, C9-B6, C9-G3, C9-E12, C9-A11, C11-E3, C11-A9 and C11-H4. PCR
15 analysis of genomic DNA isolated from the D11-C4 subclone gave rise to a 2.3 Kb band, indicative of a targeted integration into an rDNA locus. Further analysis of the subclone by FISH analysis with a DsRed gene probe demonstrated integration of the LoxDsRedLox targeting cassette on the *ACes* co-localizing with one of the regions of rDNA staining seen on
20 the 5B11-12 *ACes*, consistent with a targeted integration into an rDNA locus of the *ACes*, while integrations on other chromosomes were not observed. Since transfected cells were maintained as heterogeneous populations through several cycles of sorting and replating it was not possible to estimate the frequency of targeted events. In most
25 mammalian cell lines the frequency of gene targeting via homologous recombination is roughly 10^{-5} - 10^{-7} treated cells. Despite the low frequency of these events in mammalian cells, it is clear that an RFP expression based screening paradigm, coupled with PCR analysis, can effectively detect and enrich for such infrequent events in a large

-105-

population. In instances where drug selection is not possible or not desirable, such a system may provide a useful alternative. It was also verified that the modified *ACes* in subclone D11-C4 could be purified by flow cytometry. The results indicate that the flow karyogram of the D11-C4 subclone was unaltered from that of the 5B11-12 cell line. Thus, the D11-C4 *ACes* can be purified in high yield from native chromosomes of the host cell line.

D. Reduction of LoxP on *ACes* to a single site.

The strong hybridization signal detected by FISH on the *ACes* using the DsRed gene probe suggests that several copies of the targeting cassette may be present on the *ACes* in the D11-C4 line. This also suggests that multiple rDNA genes have been correctly targeted.

Accordingly, in certain embodiments where necessary, the number of loxP sites on the *ACes* can be reduced to a single site by *in situ* treatment with *Cre* recombinase, provided that the sites are co-linear. Such a process is described for multiple loxP-flanked integrations on a native mouse chromosome (Garrick et al., Nature Genet., 1998, Jan;18(1):56-59). Reduction to a single loxP site on the D11-C4 *ACes* would result in the loss of the DsRed gene, forming the basis of a useful screen for this event.

For this purpose, a *Cre* expression plasmid pCX-*Cre*/GFP III has been generated by first deleting the EcoRI fragment of pCX-eGFP (SEQ ID NO:71) containing the eGFP coding sequence and replacing it with that of a PCR amplified *Cre* recombinase coding sequence (SEQ ID NO:58), generating pCX-*Cre*. Next, the *Asel*/*Sspl* fragment of pD2eGFP-N1 (containing the CMV promoter driving the D2EGFP gene with SV40 polyA signal; Clontech; SEQ ID NO:87) was inserted into the filled *HindIII* site of pCX-*Cre*, generating pCX-*Cre*/GFP III. Control plasmid pCX-*Cre*Rev/GFP

-106-

III was generated in similar fashion except that the Cre recombinase coding sequence was inserted in the antisense orientation. LMTK⁻ cell line D11-C4 (containing first generation platform ACes with multiple loxP-DsRED sites) and 5B11-12 cell line (containing ACes with no loxP-DsRED sites) are maintained in culture as described above. D11C4 cells are transfected with 2 μ g of plasmid pCX-Cre\GFP III or 2 μ g pCX-CreRev\GFP III using Lipofectamine (Invitrogen) as previously described.

Forty-eight to seventy-two hours after transfection, transfected D11-C4 cells are harvested and GFP positive cells are sorted by cell cytometry using a FACSta Vantage cell sorter (Beckton-Dickinson) as follows: All D11-C4 cells transfected with pCX-Cre\GFP III or control plasmid pCX-CreRev\GFP III that exhibit GFP fluorescent higher than the gate level established by untransfected cells are collected and placed in culture a further 7-14 days. After 7-14 days the initial D11-C4 cells are harvested and analyzed by cell cytometry as follows: Untransfected D11-C4 cells are used to establish the gate that defines the RFP positive population, while 5B11-12 cells are used to set the RFP negative gate. The GFP positive population of D11-C4 transfected with pCX-Cre\GFP III should show decreased red fluorescence compared to pCX-CreRev\GFP III transfected or untransfected control D11-C4 cells. The cells exhibiting greatly decreased or no RFP expression are collected and single cell clones subsequently established. These clones will be expanded and analyzed by fluorescence *in-situ* hybridization and Southern blotting to confirm the removal of loxP-DsRed gene copies.

25

EXAMPLE 3

Construction of targeting vector and transfection into LMTK⁻ cells for the generation of platform chromosomes containing multiple site-specific recombination sites

-107-

An example of a selectable marker system for the creation of a chromosome-based platform is shown in Figure 4. This system includes a vector containing the SV40 early promoter immediately followed by (1) a 282 base pair (bp) sequence containing the bacteriophage lambda attP site and (2) the puromycin resistance marker. Initially a *PvuII/StuI* fragment containing the SV40 early promoter from plasmid pPUR (Clontech Laboratories, Inc., Palo Alto, CA; Seq ID No. 30) was subcloned into the *EcoRI/CRI* site of pNEB193 (a PUC19 derivative obtained from New England Biolabs, Beverly, MA; SEQ ID No. 32) generating the plasmid pSV40193. The only differences between pUC19 and pNEB193 are in the polylinker region. A unique *Ascl* site (GGCGCGCC) is located between the *BamHI* site and the *SmaI* site, a unique *PacI* site (TTAATTAA) is located between the *BamHI* site and the *XbaI* site and a unique *PmeI* site (GTTTAAAC) is located between the *PstI* site and the *SaI* site.

The attP site was PCR amplified from lambda genome (GenBank Accession # NC 001416) using the following primers:

attPUP: CCTTGCGCTAATGCTCTGTTACAGG SEQ ID No. 1

attPDWN: CAGAGGCAGGGAGTGGGACAAAATTG SEQ ID No. 2

After amplification and purification of the resulting fragment, the attP site was cloned into the *SmaI* site of pSV40193 and the orientation of the attP site was determined by DNA sequence analysis (plasmid pSV40193attP). The gene encoding puromycin resistance (Puro) was isolated by digesting the plasmid pPUR (Clontech Laboratories, Inc. Palo Alto, CA) with *AgeI/BamHI* followed by filling in the overhangs with Klenow and subsequently cloned into the *Ascl* site downstream of the attP site of pSV40193attP generating the plasmid pSV40193attPsensePUR (Figure 4; SEQ ID NO:113)).

-108-

The plasmid pSV40193attPsensePUR was digested with *ScaI* and co-transfected with the plasmid pFK161 (SEQ ID NO: 118) into mouse LMTk- cells and platform artificial chromosomes were identified and isolated as described above. The process for generating this exemplary platform ACes containing multiple site-specific recombination sites is summarized in Figure 5. One platform ACes resulting from this experiment is designated B19-18. This platform ACes chromosome may subsequently be engineered to contain target gene expression nucleic acids using the lambda integrase mediated site-specific recombination system as described herein in Example 7 and 8.

EXAMPLE 4

Lambda integrase mediated site-specific recombination of a RFP expressing vector onto artificial chromosomes

In this example, a vector expressing the red fluorescent protein (RFP) was produced and recombined into the attP site residing on an artificial chromosome within LMTK- cells. This recombination is depicted in Figure 7.

A. Construction of expression vectors containing wildtype and mutant lambda integrase

Mutations at the glutamic acid at position 174 in the lambda integrase protein relaxes the requirement for the accessory protein IHF during recombination and DNA supercoiling *in vitro* (see, Miller *et al.* (1980) *Cell* 20:721-729; Lange-Gustafson *et al.* (1984) *J. Biol. Chem.* 259:12724-12732). Mutations at this site promote attP, attB intramolecular recombination in mammalian cells (Lorbach *et al.* (2000) *J. Mol. Biol* 296:1175-1181).

To construct nucleic acid encoding the mutant, lambda integrase was PCR amplified from bacteriophage lambda DNA (cl857 *ind Sam* 7; New England Biolabs) using the following primers:

Lamint1 (SEQ ID No. 3)

-109-

TTCGAATTCATGGGAAGAAGGCGAAGTCATGAGCG)

Lamint2 (SEQ ID No. 4)

(TTCGAATTCTTATTTGATTTCAATTTTGTCCCAC).

- The resulting PCR product was digested with *EcoR* I and cloned into the
- 5 *EcoR* I site of pUC19. Lambda integrase was mutated at amino acid position 174 using QuikChange Site-Directed Mutagenesis Kit (Stratagene) and the following oligos (generating a glutamic acid to arginine change at position 174):

LambdaINTE174R

- 10 (SEQ ID No. 6)

(CGCGCAGCAAAATCTAGAGTAAGGAGATCAAGACTTACGGCTGACG),

LamintR174rev (SEQ ID No. 7)

(CGTCAGCCGTAAGTCTTGATCTCCTTACTCTAGATTTTGCTGCGCG).

- The resulting site directed mutant was confirmed by sequence analysis.
- 15 The wildtype and mutant lambda genes were cloned into the *EcoR* I site of pCX creating pCX-LamInt (SEQ ID NO: 127) and pCXLamIntR (Figure 8; SEQ ID NO: 112).

- The plasmid pCX (SEQ ID No. 70) was derived from plasmid pCXeGFP (SEQ ID No. 71). Excision of the *EcoRI* fragment containing the
- 20 eGFP marker generated pCX. To generate plasmid pCXLamINTR (SEQ ID NO: 112) an *EcoRI* fragment containing the lambda integrase E174R (SEQ ID No. 37) mutation was cloned into the *EcoRI* site of pCX, and to generate plasmid pCX-LamINT, an *EcoRI* fragment containing the wild-type lambda integrase was cloned into the *EcoRI* site of pCX.

- 25 **B. Construction of integration vector containing attB and DsRed**

The plasmid pDsRedN1 (Clontech Laboratories, Palo Alto, CA; SEQ ID No. 29) was digested with *Hpa* I and ligated to the following annealed oligos:

attB1 (SEQ ID No. 8)

-110-

(TGAAGCCTGCTTTTTTATACTAACTTGAGCGAA)

attB2 (SEQ ID No. 9)

(TTCGCTCAAGTTAGTATAAAAAAGCAGGCTTCA)

The resulting vector (pDsRedN1-attB) was confirmed by PCR and
5 sequence analysis.

C. Transfection into LMtk- cells

LM(tk-) cells containing the Prototype A ACes (L1-18; Chromos
Molecular Systems Inc., Burnaby, BC Canada) were co-transfected with
pDsRedN1 or pDsRedN1-attB and either pCXLamInt (SEQ ID NO: 127) or
10 pCXLamIntR (SEQ ID NO: 112) using Lipofectamine Plus Reagent
(LifeTechnologies, Gaithersburg, MD). The transfected cells were grown
in DMEM (LifeTechnologies, Gaithersburg, MD) with 10% FBS (CanSera)
and G418 (CalBiochem) at a concentration of 1 mg/ml.

D. Enrichment by cell sorting

15 The transfected cells were sorted using a FACs Vantage SE cell
sorter (Becton Dickenson) to enrich for cells expressing DsRed. The cells
were excited with a 488 nm Argon laser at 200 watts and cells
fluorescing in the 585/42 detection channel were collected. The sorted
cells were returned to growth medium for recovery and expansion. After
20 three successive enrichments for cells expressing DsRed, single cell
sorting into 96 well plates was performed using the same parameters.
Duplicate plates of the single cell clones were made for PCR analysis.

E. PCR analysis of single cell clones

25 Pools of cells from each row and column of the 96 well plate were
used for DNA isolation. DNA was prepared using a Wizard Genomic DNA
purification kit (Promega Inc, Madison, WI). Nested PCR analysis on the
DNA pools was performed to confirm the site-specific recombination
event using the following primer sets:

-111-

attPdwn2 (SEQ ID No. 10)
(TCTTCTCGGGCATAAGTCGGACACC)

CMVen (SEQ ID No. 11)
(CTCACGGGGATTTCCTCAAGTCTCCAC)

5 followed by:

attPdwn (SEQ ID No. 12)
(CAGAGGCAGGGAGTGGGACAAAATTG)

CMVen2 (SEQ ID No. 13)
(CAACTCCGCCCCATTGACGCAAATG).

- 10 The resulting PCR reactions were analyzed by gel electrophoresis and the potential individual clones containing the site-specific recombination event were identified by combining the PCR results of all of the pooled rows and columns for each 96 well plate. The individual clones were then further analyzed by PCR using the following primers that flank the
- 15 recombination junction. L1for and F1rev flank the attR junction whereas REDfor and L2rev flank the attL junction (see Figure 7):

L1for (SEQ ID No. 14)
AGTATCGCCGAACGATTAGCTCTTCA

- 20 F1rev (SEQ ID No. 15)
GCCGATTTTCGGCCTATTGGTTAAA

REDfor (SEQ ID No. 16)
CCGCCGACATCCCCGACTACAAGAA

L2rev (SEQ ID No. 17)
TTCCTTCGAAGGGGATCCGCCTACC.

25 **F. Sequence analysis of recombination junctions**

PCR products spanning the recombination junction were Topo-cloned into pcDNA3.1D/V5His (Invitrogen Inc., San Diego, CA) and then sequenced by cycle-sequencing. The clones were confirmed to have the correct *attR* and *attL* junctions by cycle sequencing.

30 **G. Fluorescent In Situ Hybridization (FISH)**

The cell lines containing the correct recombination junction sequence were further analyzed by fluorescent *in situ* hybridization (FISH)

-112-

by probing with the DsRed coding region labeled with biotin and visualizing with the Tyramide Signal Amplification system (TSA; NEN Life Science Products). The results indicate that the RFP sequence is present on the ACes.

5 **H. Southern analysis**

Genomic DNA was harvested from the cell lines containing an ACes with the correct recombinant event and digested with *EcoR* I. The digested DNAs were separated on a 0.7% agarose gel, transferred and fixed to a nylon membrane and probed with RFP coding sequences. The
10 result showed that there is an integrated copy of RFP coding sequence in each clone.

EXAMPLE 5

Delivery of a second gene encoding GFP onto the RFP platform ACes

15 **A. Construction of integration vector containing attB and GFP (pD2eGFPIresPuroattB).**

The plasmid pRESpuro2 (Clontech, Palo Alto, CA; SEQ ID NO: 88) was digested with *EcoRI* and *NotI* then ligated to the D2eGFP *EcoRI-NotI* fragment from pD2eGFP-N1 (Clontech, Palo Alto, CA) to create pD2eGFPIresPuro2. Subsequently, oligos encoding the attB site were
20 annealed and ligated into the *NruI* site of pD2eGFPIresPuro2 to create pD2eGFPIresPuroattB. The orientation of attB in the *NruI* site was determined by PCR.

B. Transfection of LMtk- cells

The LMtk- cells containing the RFP platform ACes produced in
25 Example 4, which has multiple attP sites, were co-transfected with pCXLamIntR and pD2eGFPIresPuroattB using LipofectAMINE PLUS reagent. Five μ g of each vector was placed into a tube containing 750 μ l of DMEM (Dulbecco's modified Eagles Medium). Twenty μ l of the Plus reagent was added to the DNA and incubated at room temperature for 15

-113-

minutes. A mixture of 30 μ l of lipofectamine and 750 μ l DMEM was added to the DNA mixture and incubated an additional 15 minutes at room temperature. The DNA mixture was then added dropwise to approximately 3 million cells attached to a 10cm dish in 5 mls of DMEM.

- 5 The cells were incubated 4 hours (37°C, 5% CO₂) with the DNA-lipid mixture, after which DMEM with 20% fetal bovine serum was added to the dishes to bring the culture medium to 10% fetal bovine serum. The dishes were incubated at 37°C with 5% CO₂.

Plasmid pD2eGFPIresPuroattB has a puromycin gene

- 10 transcriptionally linked to the GFP gene *via* an IRES element. Two days after the transfection the cells were placed in medium containing puromycin at 4 μ g/ml to select for cells containing the pD2eGFPIresPuroattB plasmid integrated into the genome. Twenty-three clones were isolated after 17 days of selection with puromycin. These
- 15 clones were expanded and then analyzed for the presence of the GFP gene on the ACes by 2-color (RFP/biotin & GFP/digoxigenin) TSA-FISH (NEN) according to the manufacturers protocol. Sixteen of the 23 clones produced a positive FISH signal on the ACes with a GFP probe.

EXAMPLE 6

20 Delivery Of ACes Into human Mesenchymal Stem Cells (hMSC)

A. Transfection

- Transfection conditions for the most efficient delivery of the ACes into hMSCs (Cambrex BioWhittaker Product Code PT-2501, lot# F0658, East Rutherford, New Jersey) were assayed using LipofectAMINE PLUS
- 25 and Superfect. One million prototype B ACes, which is a murine derived 60Mb ACes having primarily murine pericentric heterochromatin, and carrying a "payload" containing a hygromycin B selectable marker gene and a *lacZ* reporter gene (see , Telenius et al., 1999, Chrom. Res., 7:3-7; and Kereso et al., 1996, Chrom. Res., 4:226-239; each of which is

-114-

incorporated herein by reference in its entirety), were combined with 1-12 μ l of the transfection agent. In the case of LipofectAMINE PLUS, the PLUS reagent was combined with the ACes for 15 minutes followed by LipofectAMINE for a further 15 minutes. Superfect was complexed for 5 10 minutes at a ratio of 2 μ l Superfect per 1 million ACes. The ACes/transfection agent complex was then applied to 0.5 million recipient cells and the transfection was allowed to proceed according to the manufacturer's protocol. Percent transfected cells was determined on a FACS Vantage flow cytometer with argon laser tuned to 488 nm at 10 200mW and FITC fluorescence collected through a standard FITC 530/30 nm band pass filter. After 24 hours, IdUrd labeled ACes were delivered to human MSCs in the range of 30-50%, varying with transfection agent and dose. ACes delivery curves were generated from data collected in experiments that varied the dose of the transfection reagents. Dose 15 response curves of Superfect and LipofectAMINE PLUS, showing delivery of ACes into recipient hMSCs cells, were prepared, measured by transfer of IdUrd labeled ACes and detected by flow cytometry. Superfect shows maximum delivery in the range of 30-50% at doses greater than 2 μ l per million ACes. LipofectAMINE PLUS has a 42-48% delivery peak around 20 5-8 μ l per million ACes. These dose curves were then correlated with toxicity data to determine the transfection conditions that will allow for highest potential transfection efficiency. Toxicity was determined by a modified plating efficiency assay (de Jong et al., 2001, Chrom. Research, 9:475-485). The population's normalized plating efficiency (at maximum 25 % delivery doses) was in the range of 0.2 - 0.4 for Superfect and 0.5 - 0.6 with LipofectAMINE PLUS.

Due to the transfected population consisting of mixed cell types, flow cytometry allowed for the assessment of ACes delivery into each sub-population and the purification of the target population. Flow profiles

-115-

showing forward scatter (cell size) and side scatter (internal cell granularity) revealed three distinct hMSC populations that were gated into three regions: R3 (small cell region), R4 (medium cell region), R5 (large cell region). Transfection conditions were further optimized by re-analyzing delivery curves and assessing the differences in delivery to each sub-population. Dose response curves of Superfect and LipofectAMINE were prepared showing % delivery to each sub-population represented by the gating on basis of cell size and granularity properties of the mixed population. Three distinct hMSC populations were gated and % delivery dose curves generated. Using Superfect and LipofectAMINE PLUS the overall % delivery increased with cell size (80-90% delivery in large cells). LipofectAMINE PLUS at high doses (8-12 μ l per 1 million ACes) shows an increase in the overall proportion of chromosome transfer to the small population (10-20%). This suggests an advantage to using this transfection agent if the small-undifferentiated cell population is the desired target host cell.

B. Expression from Genes on ACes IN hMSCs

Following the delivery screening process conducted in section (A) above, the most promising results were subjected to further analyses to monitor expression and verify the presence of structurally intact ACes. The transfection conditions employed for these experiments were exactly the same as those that had been used during the screening process. Short-term expression was monitored by transfecting hMSCs with ACes containing a RFP gene (red fluorescent protein) set forth in Example 2C as "D11C4". The unselected population was harvested at 72-96 hours post transfection and % positive fluorescent cells measured by flow cytometry. RFP expression was in the range of 1-20%.

Long term-gene expression was assayed by selecting for hygromycin B resistant cells over a period of 7-10 days. Cytogenetic

-116-

analysis was done to detect presence of intact *ACes* by Fluorescent *In Situ* hybridization (FISH), where metaphase chromosomes were hybridized to a mouse major satellite-DNA probe (targeting murine pericentric heterochromatin) and a lambda probe (hybridizing to the *lacZ* gene). The human mesenchymal transfected culture could not undergo standard sub-cloning as diffuse colonies form with limited doublings available for expansion. Cytogenetic analysis was performed on the entire population, sampling over a period of 3-10 days post-transfection. The hygromycin resistant population was then blocked in mitosis with colchicine and analyzed for presence of intact *ACes* by FISH. Preliminary FISH results show approximately 2-8% of the hMSC-transfected population had an intact *ACes*. This compared to rat skeletal muscle myoblast clones, which were in the range of 60-95%. To increase the % of intact *ACes* in the hMSC-transfected population an enrichment step can be utilized as described in Example 2C.

C. Differentiation of The hMSCs

In initial experiments where transfected hMSCs cells have been induced to differentiate into adipose or osteocytes, the results indicate that the transfected cells appear to be differentiating at a rate comparable to the untransfected controls and the cultures are lineage specific as tested by microscopic examination, FISH, Oil Red O staining (adipocyte assay), and calcium secretion (osteocyte assay).

Accordingly, these results indicate that the artificial chromosomes (*ACes*) provided herein can be successfully transferred into hMSC target cells. Targeting MSCs (such as hMSCs) permits gene transfer into cells in an undifferentiated state where the cells are easier to expand and purify. The genetically modified cells can then be differentiated *in vitro* or injected into a site *in vivo* where the microenvironment will induce transformation into specific cell lineages.

-117-

EXAMPLE 7**Delivery of a Promoterless Marker Gene to a Platform ACes**

Platform ACes containing pSV40attPsensePURO (Figure 4) were constructed as set forth in Examples 3 and 4.

5 A. Construction of Targeting Vectors.

The base vector p18attBZeo (3166bp; SEQ ID NO: 114) was constructed by ligating the 1067bp *HindIII*-*SspI* fragment containing attBZeo, obtained from pLITattBZeo (SEQ ID NO:91), into pUC18 (SEQ ID NO: 122) digested with *HindIII* and *SspI*.

10 1. p18attBZEO-eGFP (6119bp; SEQ ID NO: 126) was constructed by inserting the 2977bp *SpeI*-*HindIII* fragment from pCXeGFP (SEQ ID NO:71; Okabe, *et al.* (1997) *FEBS Lett* 407:313-319) containing the eGFP gene into p18attBZeo (SEQ ID NO: 114) digested with *HindIII* and *XbaI*.

15 2. p18attBZEO-5'6XHS4eGFP (Figure 10; 7631bp; SEQ ID NO: 116) was constructed by ligating the 4465bp *HindIII* fragment from pCXeGFPattB(6XHS4)2 (SEQ ID NO: 123) which contains the eGFP gene, under the regulation of the chicken beta actin promoter, 6 copies of the HS4 core element located 5' of the chicken beta actin promoter and the polyadenylation signal into the *HindIII* site of p18attBZeo (SEQ ID NO: 20 114).

3. p18attBZEO-3'6XHS4eGFP (Figure 11; 7600bp; SEQ ID NO: 115) was created by removing the 5'6XHS4 element from p18attBZeo-(6XHS4)2eGFP (SEQ ID NO: 110). p18attBZeo-(6XHS4)2eGFP was digested with *EcoRV* and *SpeI*, treated with Klenow and religated to form 25 p18attBZeo3'6XHS4eGFP (SEQ ID NO: 115).

4. p18attBZEO-(6XHS4)2eGFP (Figure 12; 9080bp; SEQ ID NO: 110) was created in two steps. First, the *EcoRI*-*SpeI* fragment from pCXeGFPattB(6XHS4)2 (SEQ ID NO: 123) which contains 6 copies of the HS4 core element was ligated into p18attBZeo (SEQ ID NO: 114)

-118-

digested with *EcoRI* and *XbaI* to create p18attBZeo6XHS4 (4615bp; SEQ ID NO: 117). Next, p18attBZeo6XHS4 was digested with *HindIII* and ligated to the 4465bp *HindIII* fragment from pCXeGFPattB(6XHS4)2 which contains the eGFP gene, under the regulation of the chicken beta actin promoter, 6 copies of the HS4 core element located 5' of the chicken beta actin promoter and the polyadenylation signal.

Table 2

Targeting plasmid	No. zeocin resistant clones	No. clones with expected PCR product size	No. clones with correct sequence at recombination junction
p18attBZEO-eGFP	12	12	NT*
p18attBZEO-5'6XHS4eGFP	11	11	NT
p18attBZEO-3'6XHS4eGFP	11	11	NT
p18attBZEO-(6XHS4)2eGFP	9	9	4/4

*NT = not tested

B. Transfection and Selection with Drug.

The mouse cell line containing the 2nd generation platform ACE, B19-38 (constructed as set forth in Example 3), was plated onto four 10cm dishes at approximately 5 million cells per dish. The cells were incubated overnight in DMEM with 10% fetal calf serum at 37°C and 5% CO₂. The following day the cells were transfected with 5µg of each of the 4 vectors listed in Example 7.A. above and 5µg of pCXLamIntR (SEQ ID NO: 112), for a total of 10µg per 10cm dish. Lipofectamine Plus reagent was used to transfect the cells according to the manufacturers protocol. Two days post-transfection zeocin was added to the medium at 500ug/ml. The cells were maintained in selective medium until colonies formed. The colonies were then ring-cloned (see, e.g., McFarland, 2000, Methods Cell Sci, Mar;22(1):63-66).

C. Analysis of Clones (PCR, SEQUENCING).

-119-

Genomic DNA was isolated from each of the candidate clones with the Wizard kit (Promega) and following the manufacturers protocol. The following primer set was used to analyze the genomic DNA isolated from the zeocin resistant clones: 5PacSV40 -

- 5 CTGTTAATTAAGTGTGGAATGTGTG TCAGTTAGGGTG (SEQ ID NO:76);
Antisense Zeo - TGAACAGGGTCACGTCGTCC (SEQ ID NO:77). PCR amplification with the above primers and genomic DNA from the site-specific integration of any of the 4 zeocin vectors would result in a 673bp PCR product.
- 10 As set forth in Table 2, of the 4 zeocin resistant candidate clones thusfar analyzed by PCR, all 4 exhibit the correct sequence for a site-specific integration event.

EXAMPLE 8

Integration of a PCR product by site-specific recombination.

- 15 In this example a gene is integrated onto the platform *ACes* by site-specific recombination without cloning said gene into a vector.

A. PCR PRIMER DESIGN.

- PCR primers are designed to contain an attB site at the 5' end of one of the primers in the primer set. The remaining primers, which could
- 20 be one or more than one primer, do not contain an attB site, but are complementary to sequences flanking the gene or genes of interest and any associated regulatory sequences. In first example, 2 primers (one containing an attB site) are used to amplify a selective gene such as puromycin.
- 25 In a second example as shown in Figure 13, the primer set includes primers 1 & 2 that amplify the GFP gene without amplification of an upstream promoter. Primer 1 contains the attB site at the 5' end of the oligo. Primers 3 & 4 are designed to amplify the IRES-blasticidin DNA sequences from the vector pIRESblasticidin. The 5' end of primer 3

-120-

contains sequences complementary to the 5' end of primer 2 such that annealing can occur between 5' ends of the two primers.

B. PCR REACTION AND SUBSEQUENT LIGATION TO CREATE CIRCULAR MOLECULES FROM THE PCR PRODUCT

5 In the first example set forth above in Section A, the two PCR primers are combined with a puromycin DNA template such as pPUR (Clontech), a heat stable DNA polymerase and appropriate conditions for DNA amplification. The resulting PCR product (attB-Puromycin) is then then purified and self-ligated to form a circular molecule.

10 In the second example set forth above in Section A, amplification of the GFP gene and IRES-blasticidin sequences is accomplished by combining primers 1 & 2 with DNA template pD2eGFP and primers 3 & 4 with template pIRESblasticidin under appropriate conditions to amplify the desired template. After initial amplification of the two products (attB-GFP
15 & IRES-blasticidin) in separate reactions, a second round of amplification using both of the PCR products from the first round of amplification together with primers 1 and 4 amplifies the fusion product attB-GFP-IRES-blasticidin (Figure 13). This technique of using complementary sequences in primer design to create a fusion product is employed in *Saccharomyces*
20 *cerevisiae* for allele replacement (Erdeniz *et al* (1997) *Gen Res* 7:1174-1183). The amplified product is then purified from the PCR reaction mixture by standard methods and ligated to form a circular molecule.

C. INTRODUCTION OF PCR PRODUCT ONTO THE ACes USING A RECOMBINASE

25 The circular PCR product is then be introduced to the platform ACes using the bacteriophage lambda integrase E174R. The introduction can be performed *in vivo* by transfecting the pCXLamIntR (SEQ ID NO: 112) vector encoding the lambda integrase mutant E174R together with the circularized PCR product into a cell line containing the platform ACE.

-121-

D. SELECTION FOR MARKER GENE

The marker gene (in this case either puromycin, blasticidin or GFP) is used to enrich the population for cells containing the proper integration event. A proper integration event in the second example (Figure 14)

- 5 juxtaposes a promoter residing on the platform *ACes* 5' to the attB-GFP-IRES-Blasticidin PCR product, allowing for transcription of both GFP and blasticidin. If enrichment is done by drug selection, blasticidin is added to the medium on the transfected cells 24-48 hours post-transfection. Selection is maintained until colonies are formed on the plates. If
- 10 enrichment is done by cell sorting, cells are sorted 2-4 days post-transfection to enrich for cells expressing the fluorescent marker (GFP in this case).

E. ANALYSIS OF CLONES

- Clonal isolates are analyzed by PCR, FISH and sequence analysis to
- 15 confirm proper integration events.

EXAMPLE 9**Construction of a human platform *ACes* "ACE 0.1"****A. CONSTRUCTION OF THE TARGETING VECTOR pPACrDNA**

- Genome Systems (IncyteGenomics) was supplied with the primers
- 20 5'HETS (GGGCCGAAACGATCTCAACCTATT; SEQ ID NO:78), and 3'HETS (CGCAGCGGCCCTCCTACTC; SEQ ID NO:79), which were used to amplify a 538bp PCR product homologous to nt 9680-10218 of the human rDNA sequences (GenBank Accession No. U13369) and used as a probe to screen a human genomic P1AC (P1 Artificial Chromosome)
- 25 library constructed in the vector pCYPAC2 (Ioannou *et al.* (1994) *Nat. Genet.* 6(1): 84-89). Genome Systems clone #18720 was isolated in this screen and contains three repeats of human rDNA as assessed by restriction analysis. GS clone #18720, was digested with PmeI, a restriction enzyme unique to a single repeat of the human rDNA (45Kbp),

-122-

and then religated to form pPACrDNA (Figure 15). The insert in pPACrDNA was analyzed by restriction digests and sequence analysis of the 5' and 3' termini. The pPACrDNA, rDNA sequences are homologous to Genbank Accession #U13369, containing an insert of about 45 kB comprising a single repeat beginning from the end of one repeat at ~33980 (relative to the Genbank sequence) through the beginning of the next repeat up to approximately 35120 (the repeat offset from that listed in the GenBank file). Thus, the rDNA sequence is just over 1 copy of the repeat extending from 33980 (+/-10bp) to the end of the first repeat (43Kbp) and continuing into the second repeat to bp 35120 (+/-10bp).

B. TRANSFECTION AND ACes FORMATION.

Five hundred thousand MSU1.1 cells (Morgan et al., 1991, Exp. Cell Res., Nov;197(1):125-136; provided by Dr. Justin McCormick at Michigan State University) were plated per 6cm plate (3 plates total) and allowed to grow overnight. The cells were 70-80% confluent the following day. One plate was transfected with 15µg pPACrDNA (linearized with *Pme* I) and 2µg pSV40attPsensePuro (linearized with *Sca* I; see Example 3). The remaining plates were controls and were transfected with either 20µg pBS (Stratagene) or 20µg pSV40attBsensePuro (linearized with *Sca* I). All three plates were transfected using a CaPO₄ protocol.

C. SELECTION OF PUROMYCIN RESISTANT COLONIES

One day post-transfection the cells were "glycerol shocked" by the addition of PBS medium containing 10% glycerol for 30 seconds. Subsequently, the glycerol was removed and replaced with fresh DMEM. Four days post-transfection selective medium was added. Selective medium contains 1ug/ml puromycin. The transfection plates were maintained at 37°C with 5% CO₂ in selective medium for 2 weeks at which point colonies could be seen on the plate transfected with

-123-

pPACrDNA and pSV40attPsensePuro. The colonies were ring-cloned from the plate on day 17 post-selection and expanded in selective medium for analysis. Only two colonies (M2-2d & M2-2b) were able to proliferate in the selective medium after cloning. No colonies were seen on the control plates after 37 days in selective medium.

D. ANALYSIS OF CLONES

FISH analysis was performed on the candidate clones to detect ACes formation. Metaphase spreads from the candidate clones were probed in multiple probe combinations. In one experiment, the probes used were biotin-labeled human aliphoid DNA (pPACrDNA) and digoxigenin-labeled mouse major DNA (pFK161) as a negative control. Candidate M2-2d was single cell subcloned by flow sorting and the candidate subclones were reanalyzed by FISH. Subclone 1B1 of M2-2d was determined to be a platform ACes and is also designated human Platform ACE 0.1.

EXAMPLE 10

Site-specific integration of a marker gene onto a human platform ACE 0.1

The promoterless delivery method was used to deliver a promoterless blasticidin marker gene onto the human platform ACes with excellent results. The human ACes platform with a promoterless blasticidin marker gene resulted in 21 of 38 blasticidin resistant clones displaying a PCR product of the expected size from the population co-transfected with pLIT38attBBSRpolyA10 and pCXLamIntR (Figure 8; SEQ ID NOs. 111 and 112). Whereas, the population transfected with pBlueScript resulted in 0 blasticidin resistant colonies.

A. CONSTRUCTION OF pLIT38attB-BSRpolyA10 & pLIT38attB-BSRpolyA2.

The vector pLITMUS 38 (New England Biolabs; U.S. Patent No. 5,691,140; SEQ ID NO: 119) was digested with *EcoRV* and ligated to

-124-

two annealed oligomers, which form an attB site (attB1 5'-TGAAGCCTGCTTTTTTATACTAACTTGAGCGAA-3' (SEQ ID NO:8); attB2 5'- TTCGCTCAAGTTAGTATAAAAAAGCAGGCTTCA-3'; SEQ ID NO:9). This ligation reaction resulted in the vector pLIT38attB (SEQ ID NO: 120).

- 5 The blasticidin resistance gene and SV40 polyA site was PCR amplified with primers: 5BSD (ACCATGAAAACATTTAACATTTCTCAACA; SEQ ID NO:80) and SV40polyA (TTTATTTGTGAAATTTGTGATGCTATTGC; SEQ ID NO:81) using pPAC4 (Frengen, E., *et al.* (2000) Genomics 68 (2), 118-126; GenBank Accession No. U75992) as template. The blasticidin-
- 10 SV40polyA PCR product was then ligated into pLIT38attB at the *Bam*HI site, which was Klenow treated following digestion with *Bam*HI. pLIT38attB-BSDpolyA10 (SEQ ID NO: 111) and pLIT38attB-BSDpolyA2 (SEQ ID NO: 121) are the two resulting orientations of the PCR product ligated into the vector.

15 **B. TRANSFECTION OF MSU1.1 CELLS CONTAINING HUMAN PLATFORM ACE 0.1.**

- MSU1.1 cells containing human platform ACE 0.1 (see Example 9) was expanded and plated to five 10cm dishes with 1.3×10^6 cells per dish. The cells were incubated overnight in DMEM with 10% fetal bovine
- 20 serum, at 37°C and 5% CO₂. The following day the cells were transfected with 5µg of each plasmid as set forth in Table 3, for a total of 10µg of DNA per plate of cells transfected (see Table 3) using ExGen 500 *in vitro* transfection reagent (MBI fermentas, cat. no. R0511). The transfection was performed according to the manufacturers protocol.
- 25 Cells were incubated at 37°C with 5% CO₂ in DMEM with 10% fetal bovine serum following the transfection.

-125-

Table 3

Plate #	Plasmid 1	Plasmid 2	No. Bsd ^R Colonies
1	pBS	None	0
2	pCXLamInt	pLIT38attB-BSRpolyA10	16
3	pCXLamIntR	pLIT38attB-BSRpolyA10	40
4	pCXLamInt	pLIT38attB-BSRpolyA2	28
5	pCXLamIntR	pLIT38attB-BSRpolyA2	36

10 C. SELECTION OF BLASTICIDIN RESISTANT CLONES.

Three days following the transfection the cells were split from a 10 cm dish to two 15cm dishes. The cells were maintained in DMEM with 10% fetal bovine serum for 4 days in the 15 cm dishes. Seven days post-transfection blasticidin was introduced into the medium. Stably transfected cells were selected with 1 μ g/ml blasticidin. The number of colonies formed on each plate is listed in Table 3. These colonies were ring-cloned and expanded for PCR analysis. Upon expansion in blasticidin containing medium some clones failed to live and therefore do not have corresponding PCR data.

20 D. PCR ANALYSIS

Thirty-eight of the 40 clones from plate 3 grew after ring-cloning. Genomic DNA was isolated from these clones with the Promega Wizard Genomic cDNA purification kit, digested with *Eco*RI and used as template in a PCR reaction with the following primers: 3BSP – TTAATTTCGGG TATATTTGAGTGGA (SEQ ID NO:82); 5PacSV40 – CTGTTAATTAAGTGTGGAA TGTGTGTCAGTTAGGGTG (SEQ ID NO:76). The PCR conditions were as follows. 100ng of genomic DNA was

-126-

amplified with 0.5ul Herculase polymerase (Stratagene) in a 50ul reaction that contained 12.5pmole of each primer, 2.5mM of each dNTP, and 1X Herculase buffer (Stratagene). The reactions were placed in a PerkinElmer thermocycler programmed as follows: Initial denaturation at 95°C for 10 minutes; 35 cycles of 94°C for 1 minute, 53°C for 1 minute, 72°C for 1 minute, and 72°C for 1 minute; Final extension for 10 minutes at 72°C; and 4°C hold. If pLIT38attB-BSRpolyA10 integrates onto the human platform ACE 0.1 correctly, PCR amplification with the above primers should yield an 804bp product. Twenty-one of the 38 clones from plate 3 produced a PCR product of the expected 804bp size.

EXAMPLE 11

Delivery of a Vector comprising a Promoterless Marker Gene and a gene encoding a therapeutic product to a Platform ACes

Platform ACes containing pSV40attPsensePURO (Figure 4) were constructed as set forth in Examples 3 and 4.

A. CONSTRUCTION OF DELIVERY VECTORS

1. Erythropoietin cDNA vector, p18EPOcDNA.

The erythropoietin cDNA was PCR amplified from a human cDNA library (E. Perkins *et al.*, 1999, *Proc. Natl. Acad. Sci. USA* 96(5): 2204-2209) using the following primers: EPO5XBA - TATCTAGAATGGGGGTGC ACGAATGTCCTGCC (SEQ ID NO: 83); EPO3BSI - TACGTACGTCATC TGTCCTGTCCTGCAGGC (SEQ ID NO: 84). The cDNA was amplified through two successive rounds of PCR using the following conditions: heat denaturation at 95°C for 3 minutes; 35 cycles of a 30 second denaturation (95°C), 30 seconds of annealing (60°C), and 1 minute extension (72°C); the last cycle is followed by a 7 minute extension at 72°C. BIO-X-ACT (BIOLINE) was used to amplify the erythropoietin cDNA from 2.5ng of the human cDNA library in the first round of amplification. Five μ l of the first amplification product was used

-127-

as template for the second round of amplification. Two PCR products were produced from the second amplification with Taq polymerase (Eppendorf), each product was cloned into pCR2.1-Topo (Invitrogen) and sequenced. The larger PCR product contained the expected cDNA
5 sequence for erythropoietin. The erythropoietin cDNA was moved from pTopoEPO into p18attBZeo(6XHS4)2eGFP (SEQ ID NO: 110). pTopoEPO was digested with BsiWI and XbaI to release a 588 bp EPO cDNA. BsrGI and BsiWI create compatible ends. The eGFP gene was removed from p18attBZeo(6XHS4)2eGFP by digestion with BsiWI and XbaI, the 8.3 Kbp
10 vector backbone was gel purified and ligated to the 588 bp EPO cDNA to create p18EPOcDNA (SEQ ID NO: 124).

2. Genomic erythropoietin vector, p18genEPO.

The erythropoietin genomic clone was PCR amplified from a human genomic library (Clontech) using the following primers: GENEPO3BSI -
15 CGTACGTCATCTGTCCCCT GTCCTGCA (SEQ ID NO: 85); GENEPO 5XBA -TCTAGAATGGGGGT GCACGGTGAGTACT (SEQ ID NO: 86). The reaction conditions for the amplification were as follows: heat denaturation for 3 minutes (95°C); 30 cycles of a 30 second denaturation (95°C), 30 seconds annealing (from 65°C decreasing 0.5°C per cycle to
20 50°C), and 3 minutes extension (72°C); 15 cycles of a 30 second denaturation (95°C), 30 seconds annealing (50°C), and 3 minute extension (72°C); the last cycle is followed by a 7 minute extension at 72°C. The erythropoietin genomic PCR product (2147 bp) was gel purified and cloned into pCR2.1Topo to create pTopogenEPO. Sequence
25 analysis revealed 2bp substitutions and insertions in the intronic sequences of the genomic clone of erythropoietin. A partial digest with XbaI and complete digest with BsiWI excised the erythropoietin genomic insert from pTopogenEPO. The resulting 2158 bp genomic erythropoietin fragment was ligated into the 8.3 Kbp fragment resulting from the

-128-

digestion of p18attBZeo(6XHS4)2eGFP (SEQ ID NO: 110) with XbaI and BsrGI to create p18genEPO (SEQ ID NO: 125).

B. TRANSFECTION AND SELECTION WITH DRUG

The erythropoietin genomic and cDNA genes were each moved
5 onto the platform ACes B19-38 (constructed as set forth in Example 3) by
co-transfecting with pCXLamIntR. Control transfections were also
performed using pCXLamInt (SEQ ID NO: 127) together with either
p18EPOcDNA (SEQ ID NO: 124) or p18genEPO (SEQ ID NO: 125).
Lipofectamine Plus was used to transfect the DNA's into B19-38 cells
10 according to the manufacturer's protocol. The cells were placed in
selective medium (DMEM with 10% FBS and Zeocin @ 500ug/ml) 48
hours post-transfection and maintained in selective medium for 13 days.
Clones were isolated 15 days post-transfection.

C. ANALYSIS OF CLONES (ELISA, PCR)

15 1. ELISA Assays

Thirty clones were tested for erythropoietin production by an ELISA
assay using a monoclonal anti-human erythropoietin antibody (R&D
Systems, Catalogue # MAB287), a polyclonal anti-human erythropoietin
antibody (R & D Systems, Catalogue # AB-286-NA) and alkaline
20 phosphatase conjugated goat-anti-rabbit IgG (heavy and light chains)
(Jackson ImmunoResearch Laboratories, Inc., Catalogue # 111-055-144).
The negative control was a Zeocin resistant clone isolated from B19-38
cells transfected with p18attBZeo(6XHS4) (SEQ ID NO: 117; no insert
control vector) and pCXLamIntR (SEQ ID NO: 112). The preliminary
25 ELISA assay was executed as follows: 1) Nunc-Immuno Plates (MaxiSorb
96-well, Catalogue # 439454) were coated with 75ul of a 1/200 dilution
(in Phosphate buffered Saline, pH 7.4 (PBS), Sigma Catalogue # P-3813)
of monoclonal anti-human erythropoietin antibody overnight at 4°C. 2)
The following day the plates were washed 3 times with 300ul PBS

-129-

containing 0.15% Tween 20 (Sigma, Catalogue # P-9416). 3) The plates were then blocked with 300ul of 1% Bovine Serum Albumin (BSA; Sigma Catalogue # A-7030) in PBS for 1 hour at 37°C. 4) Repeat the washes as in step 2. 5) The clonal supernatants (75ul per clone per well of 96-well plate) were then added to the plate and incubated for 1 hour at 37°C. The clonal supernatant analyzed in the ELISA assay had been maintained on the cells 7 days prior to analysis. 6) Repeat the washes of step 2. 7) Add 75ul of polyclonal anti-human erythropoietin antibody (1/250 dilution in dilution buffer (0.5% BSA, 0.01% Tween 20, 1X PBS, pH 7.4) and incubate 1 hour at 37°C. 8) Repeat washes of step 2. 9) Add 75ul of goat anti-rabbit conjugated alkaline phosphatase diluted 1/4000 in dilution buffer and incubate 1 hour at 37°C. 10) Repeat washes of step 2. 11) Add 75ul substrate, p-nitrophenyl phosphate (Sigma N2640), diluted to 1mg/ml in substrate buffer (0.1 Ethanolamine-HCl (Sigma, Catalogue # E-6133), 5mM MgCl₂ (Sigma, Catalogue # M-2393), pH 9.8). Incubate the plates in the dark for 1 hour at room temperature (22°C). 12) Read the absorption at 405nm (reference wavelength 495nm) on an Universal Microplate Reader (Bio-Tek Instruments, Inc., model # ELX800 UV). The erythropoietin standard curve was derived from readings of diluted human recombinant Erythropoietin (Roche, catalogue # 1-120-166; dilution range 125 - 7.8mUnits/ml). From this preliminary assay the 21 clones displaying the highest expression of erythropoietin were analyzed a second time in the same manner using medium supernatants that had been on the clones for 24 hours and a 1:3 dilution thereof.

25 2. PCR Analysis

Genomic DNA was isolated from the 21 clones with the best expression (as assessed by the initial ELISA assay above) as well as the B19-38 cell line and used for PCR analysis. Genomic DNA was isolated using the Wizard genomic DNA purification kit (Promega) according to the

-130-

manufacturers protocol. Amplification was performed on 100ng of genomic DNA as template with MasterTaq DNA Polymerase (Eppendorf) and the primer set 5PacSV40 – CTGTTAATTAAGTGTGGAATGTGTG TCAGTTAGGGTG (SEQ ID NO: 76) and Antisense Zeo -

- 5 TGAACAGGGTCACGTCGTCC (SEQ ID NO: 77). The amplification conditions were as follows: heat denaturation for 3 minutes (95°C); 30 cycles of a 30 second denaturation (95°C), 30 seconds annealing (from 65°C decreasing 0.5°C per cycle to 50°C), and 1 minutes extension (72°C); 15 cycles of a 30 second denaturation (95°C), 30 seconds
- 10 annealing (50°C), and 1 minute extension (72°C); the last cycle is followed by a 10 minute extension at 72°C. PCR products were size separated by gel electrophoresis. Of the 21 clones analyzed 19 produced a PCR product of 650 bp as expected for a site-specific integration event. All nineteen clones were the result of transformations with p19EPOcDNA
- 15 (5) or p18genEPO (14) and pCXLamIntR (i.e. mutant integrase). The remaining two clones, both of which were the result of transformation with p18genEPO (SEQ ID NO: 125) and pCXLamInt (i.e. wildtype integrase; SEQ ID NO: 127), produced a 400 bp PCR product.

Example 12

20 Preparation of a Transformation Vector Useful for the Induction of Plant Artificial Chromosome Formation

- Plant artificial chromosomes (PACs) can be generated by introducing nucleic acid, such as DNA, which can include a targeting DNA, for example rDNA or lambda DNA, into a plant cell, allowing the cell
- 25 to grow, and then identifying from among the resulting cells those that include a chromosome with a structure that is distinct from that of any chromosome that existed in the cell prior to introduction of the nucleic acid. The structure of a PAC reflects amplification of chromosomal DNA, for example, segmented, repeat region-containing and heterochromatic

-131-

structures. It is also possible to select cells that contain structures that are precursors to PACs, for example, chromosomes containing more than one centromere and/or fragments thereof, and culture and/or manipulate them to ultimately generate a PAC within the cell.

5 In the method of generating PACs, the nucleic acid can be introduced into a variety of plant cells. The nucleic acid can include targeting DNA and/or a plant expressible DNA encoding one or multiple selectable markers (*e.g.*, DNA encoding bialaphos (bar) resistance) or scorable markers (*e.g.*, DNA encoding GFP). Examples of targeting DNA
10 include, but are not limited to, *N. tabacum* rDNA intergenic spacer sequence (IGS) and *Arabidopsis* rDNA such as the 18S, 5.8S, 26S rDNA and/or the intergenic spacer sequence. The DNA can be introduced using a variety of methods, including, but not limited to *Agrobacterium*-mediated methods, PEG-mediated DNA uptake and electroporation using,
15 for example, standard procedures according to Hartmann *et al* [(1998) *Plant Molecular Biology* 36:741]. The cell into which such DNA is introduced can be grown under selective conditions and can initially be grown under non-selective conditions and then transferred to selective media. The cells or protoplasts can be placed on plates containing a
20 selection agent to grow, for example, individual calli. Resistant calli can be scored for scorable marker expression. Metaphase spreads of resistance cultures can be prepared, and the metaphase chromosomes examined by FISH analysis using specific probes in order to detect amplification of regions of the chromosomes. Cells that have artificial
25 chromosomes with functioning centromeres or artificial chromosomal intermediate structures, including, but not limited to, dicentric chromosomes, formerly dicentric chromosomes, minichromosomes, heterochromatin structures (*e.g.* sausage chromosomes), and stable self-replicating artificial chromosomal intermediates as described herein, are

-132-

identified and cultured. In particular, the cells containing self-replicating artificial chromosomes are identified.

The DNA introduced into a plant cell for the generation of PACs can be in any form, including in the form of a vector. An exemplary
5 vector for use in methods of generating PACs can be prepared as follows.

For the production of artificial chromosomes, plant transformation vectors, as exemplified by pAgIIa and pAgIIb, containing a selectable marker, a targeting sequence, and a scorable marker were constructed using procedures well known in the art to combine the various fragments.
10 The vectors can be prepared using vector pAg1 as a base vector and inserting the following DNA fragments into pAg1: DNA encoding β -glucoronidase under the control of the nopaline synthase (NOS) promoter fragment and flanked at the 3' end by the NOS terminator fragment, a fragment of mouse satellite DNA and an *N. tabacum* rDNA intergenic
15 spacer sequence (IGS). In constructing plant transformation vectors, vector pAg2 can also be used as the base vector.

1. Construction of pAG1

Vector pAg1 (SEQ. ID. NO: 89) is a derivative of the CAMBIA vector named pCambia 3300 (Center for the Application of Molecular
20 Biology to International Agriculture, i.e., CAMBIA, Canberra, Australia; www.cambia.org), which is a modified version of vector pCambia 1300 to which has been added DNA from the bar gene conferring resistance to phosphinothricin. The nucleotide sequence of pCambia 3300 is provided in SEQ. ID. NO: 90. pCambia 3300 also contains a lacZ alpha sequence
25 containing a polylinker region.

pAg1 was constructed by inserting two new functional DNA fragments into the polylinker of pCambia 3300: one sequence containing an attB site and a promoterless zeomycin resistance-encoding DNA flanked at the 3' end by a SV40 polyA signal sequence, and a second

-133-

sequence containing DNA from the hygromycin resistance gene (hygromycin phosphotransferase) conferring resistance to hygromycin for selection in plants. Although the zeomycin-SV40 polyA signal fusion is not expected to function in plant cells, it can be activated in mammalian
5 cells by insertion of a functional promoter element into the attB site by site-specific recombination catalyzed by the Lambda att integrase. Thus, the inclusion of the attB-zeomycin sequences allows for evaluation of functionality of plant artificial chromosomes in mammalian cells by activation of the zeomycin resistance-encoding DNA, and provides an att
10 site for further insertion of new DNA sequences into plant artificial chromosomes formed as a result of using pAg1 for plant transformation. The second functional DNA fragment allows for selection of plant cells with hygromycin. Thus, pAg1 contains DNA from the bar gene conferring resistance to phosphinothricin, DNA from the hygromycin resistance gene,
15 both resistance-encoding DNAs under the control of a separate cauliflower mosaic virus (CaMV) 35S promoter, and the attB-promoterless zeomycin resistance-encoding DNA.

pAg1 is a binary vector containing *Agrobacterium* right and left T-DNA border sequences for use in *Agrobacterium*-mediated transformation
20 of plant cells or protoplasts with the DNA located between the border sequences. pAg1 also contains the pBR322 Ori for replication in *E. coli*. pAg1 was constructed by ligating *Hind*III/*Pst*I-digested p3300attBZeo with *Hind*III/*Pst*I-digested pBSCaMV35SHyg as follows.

a. **Generation of p3300attBZeo**

25 Plasmid pCambia 3300 was digested with *Pst*I/*Ec*1136 II and ligated with *Pst*I/*Stu*I-digested pLITattBZeo (the nucleotide sequence of pLITattBZeo is provided in SEQ. ID. NO: 91. (containing DNA encoding the zeocin resistance gene and an attB Integrase recognition sequence) to generate p3300attBZeo which contains an attB site, a promoterless

-134-

zeomycin resistance-encoding DNA flanked at the 3' end by a SV40 polyA signal, and a reconstructed *Pst*I site.

b. Generation of pBSCaMV35SHyg

A DNA fragment containing DNA encoding hygromycin

5 phosphotransferase flanked by the CaMV 35S promoter and the CaMV 35S polyA signal sequence was obtained by PCR amplification of plasmid pCambia 1302 (GenBank Accession No. AF234298 and SEQ. ID. NO: 92). The primers used in the amplification reaction were as follows:
CaMV35SpolyA:

10 5'-CTGAATTAACGCCGAATTAATTCGGGGGATCTG-3' SEQ. ID. NO: 93
CaMV35Spr:

5'-CTAGAGCAGCTTGCCAACATGGTGGAGCA-3' SEQ. ID. NO: 94

The 2100-bp PCR fragment was ligated with *Eco*RV-digested pBluescript II SK+ (Stratagene, La Jolla, CA, U.S.A.) to generate pBSCaMV35SHyg.

15 **c. Generation of pAg1**

To generate pAg1, pBSCaMV35SHyg was digested with *Hind*III/*Pst*I and ligated with *Hind*III/*Pst*I-digested p3300attBZeo. Thus, pAg1 contains the pCambia 3300 backbone with DNA conferring resistance to phophinothricin and hygromycin under the control of
20 separate CaMV 35S promoters, an attB-promoterless zeomycin resistance-encoding DNA recombination cassette and unique sites for adding additional markers, *e.g.*, DNA encoding GFP. The attB site can be used as described hereing for the addition of new DNA sequences to plant artificial chromosomes, including PACs formed as a result of using the
25 pAg1 vector, or derivatives thereof, in the production of PACs. The attB site provides a convenient site for recombinase-mediated insertion of DNAs containing a homologous att site.

2. pAG2

-135-

The vector pAg2 (SEQ. ID. NO: 95) is a derivative of vector pAg1 formed by adding DNA encoding a green fluorescent protein (GFP), under the control of a NOS promoter and flanked at the 3' end by a NOS polyA signal, to pAg1. pAg2 was constructed as follows. A DNA fragment
5 containing the NOS promoter was obtained by digestion of pGEM-T-NOS, or pGEMEasyNOS (SEQ. ID. NO: 96), containing the NOS promoter in the cloning vector pGEM-T-Easy (Promega Biotech, Madison, WI, U.S.A.), with *Xba*I/*Nco*I and was ligated to an *Xba*I/*Nco*I fragment of pCambia 1302 containing DNA encoding GFP (without the CaMV 35S promoter) to
10 generate p1302NOS (SEQ. ID. NO: 97) containing GFP-encoding DNA in operable association with the NOS promoter. Plasmid p1302NOS was digested with *Sma*I/*Bsi*WI to yield a fragment containing the NOS promoter and GFP-encoding DNA. The fragment was ligated with *Pme*I/*Bsi*WI-digested pAg1 to generate pAg2. Thus, pAg2 contains DNA
15 from the bar gene conferring resistance to phosphinothricin, DNA conferring resistance to hygromycin, both resistance-encoding DNAs under the control of a cauliflower mosaic virus 35S promoter, DNA encoding kanamycin resistance, a GFP gene under the control of a NOS promoter and the attB-zeomycin resistance-encoding DNA. One of skill in
20 the art will appreciate that other fragments can be used to generate the pAg1 and pAg2 derivatives and that other heterologous DNA can be incorporated into pAg1 and pAg2 derivatives using methods well known in the art.

3. pAgIIa and pAgIIb transformation vectors

25 Vectors pAgIIa and pAgIIb were constructed by inserting the following DNA fragments into pAg1: DNA encoding β -glucuronidase, the nopaline synthase terminator fragment, the nopaline synthase (NOS) promoter fragment, a fragment of mouse satellite DNA and an *N. tabacum*

-136-

rDNA intergenic spacer sequence (IGS). The construction of pAgIIa and pAgIIb was as follows.

An *N. tabacum* rDNA intergenic spacer (IGS) sequence (SEQ. ID. NO: 98; see also GenBank Accession No. YO8422; see also Borysyuk *et al.* (2000) *Nature Biotechnology* 18:1303-1306; Borysyuk *et al.* (1997) *Plant Mol. Biol.* 35:655-660; U.S. Patent Nos. 6,100,092 and 6,355,860) was obtained by PCR amplification of tobacco genomic DNA. The IGS can be used as a targeting sequence by virtue of its homology to tobacco rDNA genes; the sequence is also an amplification promoter sequence in plants. This fragment was amplified using standard PCR conditions (*e.g.*, as described by Promega Biotech, Madison, WI, U.S.A.) from tobacco genomic DNA using the primers shown below:

NTIGS-FI

5'- GTG CTA GCC AAT GTT TAA CAA GAT G- 3' (SEQ ID No. 99) and

NTIGS-RI

5'-ATG TCT TAA AAA AAA AAA CCC AAG TGA C- 3' (SEQ ID No. 100)

Following amplification, the fragment was cloned into pGEM-T Easy to give pIGS-I. A fragment of mouse satellite DNA (Msat1 fragment; GenBank Accession No. V00846; and SEQ ID No. 101) was amplified via

PCR from pSAT-1 using the following primers:

MSAT-F1

5'- AAT ACC GCG GAA GCT TGA CCT GGA ATA TCG C -3' (SEQ ID No. 102) and

MSAT-RI

5'-ATA ACC GCG GAG TCC TTC AGT GTG CA T- 3' (SEQ ID No. 103)

This amplification added a *SacII* and a *HindIII* site at the 5' end and a *SacII* site at the 3' end of the PCR fragment. This fragment was then cloned into the *SacII* site in pIGS-1 to give pMIGS-1, providing a eukaryotic

-137-

centromere-specific DNA and a convenient DNA sequence for detection via FISH.

A functional marker gene containing a NOS-promoter:GUS:NOS terminator fusion was then constructed containing the NOS promoter
5 (GenBank Accession No. U09365; SEQ ID No. 104), *E. coli*
 β -glucuronidase coding sequence (from the GUS gene; GenBank
Accession No. S69414; and SEQ ID No. 105), and the nopaline synthase
terminator sequence (GenBank Accession No. U09365; SEQ ID No. 107).
The NOS promoter in pGEM-T-NOS was added to a promoterless GUS
10 gene in pBlueScript (Stratagene, La Jolla, CA, U.S.A.) using *NotI*/*SpeI* to
form pNGN-1, which has the NOS promoter in the opposite orientation
relative to the GUS gene.

pMIGS-1 was digested with *NotI*/*SpeI* to yield a fragment
containing the mouse major satellite DNA and the tobacco IGS which was
15 then added to *NotI*-digested pNGN-1 to yield pNGN-2. The NOS promoter
was then re-oriented to provide a functional GUS gene, yielding pNGN-3,
by digestion and religation with *SpeI*. Plasmid pNGN-3 was then digested
with *HindIII*, and the *HindIII* fragment containing the β -glucuronidase
coding sequence and the rDNA intergenic spacer, along with the Msat
20 sequence, was added to pAG-1 to form pAgIIa (SEQ ID NO: 108), using
the unique *HindIII* site in pAg1 located near the right T-DNA border of
pAg1, within the T-DNA region.

Another plasmid vector, referred to as pAgIIb, was also recovered,
which contained the inserted *HindIII* fragment (SEQ ID NO: 108) in the
25 opposite orientation relative to that observed in pAgIIa. Thus, pAgIIa and
pAgIIb differ only in the orientation of the *HindIII* fragment containing the
mouse major satellite sequence, the GUS DNA sequence and the IGS
sequence. The nucleotide sequences of pAgIIa is provided in SEQ. ID.
NOS: 109.

-138-

Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

-139-

WHAT IS CLAIMED IS:

1. A eukaryotic chromosome comprising one or a plurality of *att* site(s), wherein:
an *att* site is heterologous to the chromosome; and
5 an *att* site permits site-directed integration in the presence of lambda integrase.
2. The eukaryotic chromosome of claim 1, wherein the *att* sites are selected from the group consisting of *attP* and *attB* or *attL* and *attR*, or variants thereof.
- 10 3. The eukaryotic chromosome of claim 1 that is an artificial chromosome.
4. The eukaryotic chromosome of claim 1 that is an artificial chromosome expression system (*ACes*).
5. The eukaryotic chromosome of claim 4 that is predominantly
15 heterochromatin.
6. The chromosome of claim 1 that is an artificial chromosome that contains no more than about 30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% euchromatin.
7. The chromosome of claim 1 that is a plant chromosome.
- 20 8. The chromosome of claim 1 that is an animal chromosome.
9. The chromosome of claim 7 that is a plant artificial chromosome.
10. The chromosome of claim 8 that is an animal artificial chromosome.
- 25 11. The chromosome of claim 8 that is a mammalian chromosome.
12. The chromosome of claim 11 that is a mammalian artificial chromosome.

-140-

13. The chromosome of claim 6 that is an artificial chromosome expression system (*ACes*).

14. A platform artificial chromosome expression system (*ACes*) comprising one or a plurality of sites that participate in recombinase
5 catalyzed recombination.

15. The *ACes* of claim 14 that contains one site.

16. The *ACes* of claim 14 that is predominantly heterochromatin.

17. The *ACes* of claim 14 that contains no more than about 30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% euchromatin.

10 18. The *ACes* of claim 14 that is a plant *ACes*.

19. The *ACes* of claim 14 that is an animal *ACes*.

20. The *ACes* of claim 14 that is selected from a fish, insect, reptile, amphibian, arachnid or a mammalian *ACes*.

21. The *ACes* of claim 14 that is a fish *ACes*.

15 22. The artificial chromosome expression system (*ACes*) of claim 14, wherein the recombinase and site(s) are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase system of phage Mu, the Cin recombinase system, the Pin recombinase system of *E. coli* and the R/RS
20 system of the pSR1 plasmid, or any combination thereof.

23. A method of introducing heterologous nucleic acid into a chromosome, comprising:

contacting a chromosome of any of claims 1 or 14 with a nucleic acid molecule comprising both the heterologous nucleic acid and a
25 recombination site, in the presence of a recombinase that promotes recombination between the sites in the chromosome and in the nucleic acid molecule.

-141-

24. The method of claim 23, wherein the recombinase is selected from the group consisting of Cre, Gin, Cin, Pin, FLP, a phage integrase and R from the pSR1 plasmid.

25. The method of claim 23, wherein the nucleic acid molecule
5 encodes a therapeutic protein, antisense nucleic acid, or comprises an artificial chromosome.

26. The method of claim 25, wherein the nucleic acid molecule comprises a yeast artificial chromosomes (YAC), a bacterial artificial chromosome (BAC) or an insect artificial chromosome (IAC).

10 27. A combination, comprising, the chromosome of claim 1 and a first vector comprising the cognate recombination site, wherein the cognate recombination site is a site that recombines with the site engineered into the chromosome.

28. The combination of claim 27, further comprising nucleic acid
15 encoding a recombinase, wherein the nucleic acid is on a second vector or on the first vector, or on the ACes under an inducible promoter.

29. The combination of claim 28, wherein the recombinase and sites are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase
20 system of phage Mu, the Pin recombinase system of *E. coli* and the R/RS system of the pSR1 plasmid, or any combination thereof.

30. The combination of claim 28, wherein a vector is the plasmid pCXLamIntR.

31. The combination of claim 27, wherein a vector is the plasmid
25 pDsRedN1-attB.

32. A kit, comprising the combination of claim 27 and optionally instructions for introducing heterologous nucleic acid into the chromosome.

-142-

33. A method for introducing heterologous nucleic acid into a platform artificial chromosome, comprising:

(a) mixing an artificial chromosome comprising at least a first recombination site and a vector comprising at least a second

5 recombination site and the heterologous nucleic acid;

(b) incubating the resulting mixture in the presence of at least one recombination protein under conditions whereby recombination between the first and second recombination sites is effected, thereby introducing the heterologous nucleic acid into the artificial chromosome.

10 34. The method of claim 33, wherein the artificial chromosome is an ACes.

35. The method of claim 33, wherein said mixing step (a) is conducted in cells ex vivo.

15 36. The method of claim 33, wherein said mixing step (a) is conducted extracellularly in an in vitro reaction mixture.

37. The method of claim 33, wherein the at least one recombination protein is encoded by a bacteriophage selected from the group consisting of bacteriophage lambda, phi 80, P22, P2, 186, P4 and P1.

20 38. The method of claim 37, wherein the at least one recombination protein is encoded by bacteriophage lambda, or mutants thereof.

39. The method of claim 33, wherein at least one recombination protein is selected from the group consisting of Int, IHF, Xis and Cre, $\gamma\delta$,
25 Tn3 resolvase, Hin, Gin, Cin and Flp.

40. The method of claim 32, wherein the recombination sites are selected from the group consisting of att and lox P sites.

-143-

41. The method of claim 33, wherein the first and/or second recombination site contains at least one mutation that removes one or more stop codons.

42. The method of claim 33, wherein the first and/or second
5 recombination site contains at least one mutation that avoids hairpin formation.

43. The method of claim 33, wherein the first and/or second recombination site comprises at least a first nucleic acid sequence selected from the group consisting of SEQ ID NOs:41-56:

- 10 a) RKYCWGCTTTYKTRTACNAASTSGB (m-att) (SEQ ID NO:41);
b) AGCCWGCTTTYKTRTACNAACTSGB (m-attB) (SEQ ID NO:42);
c) GTTCAGCTTTCKTRTACNAACTSGB (m-attR) (SEQ ID NO:43);
d) AGCCWGCTTTCKTRTACNAAAGTSGB (m-attL) (SEQ ID NO:44);
e) GTTCAGCTTTYKTRTACNAAAGTSGB (m-attP1) (SEQ ID NO:45);
15 f) AGCCTGCTTTTTTGTACAAACTTGT (attB1) (SEQ ID NO:46);
g) AGCCTGCTTTCTTGTACAAACTTGT (attB2) (SEQ ID NO:47);
h) ACCCAGCTTTCTTGTACAAACTTGT (attB3) (SEQ ID NO:48);
i) GTTCAGCTTTTTTGTACAAACTTGT (attR1) (SEQ ID NO:49);
j) GTTCAGCTTTCTTGTACAAACTTGT (attR2) (SEQ ID NO:50);
20 k) GTTCAGCTTTCTTGTACAAAGTTGG (attR3) (SEQ ID NO:51);
l) AGCCTGCTTTTTTGTACAAAGTTGG (attL1) (SEQ ID NO:52);
m) AGCCTGCTTTCTTGTACAAAGTTGG (attL2) (SEQ ID NO:53);
n) ACCCAGCTTTCTTGTACAAAGTTGG (attL3) (SEQ ID NO:54);
o) GTTCAGCTTTTTTGTACAAAGTTGG (attP1) (SEQ ID NO:55);
25 p) GTTCAGCTTTCTTGTACAAAGTTGG (attP2, P3) (SEQ ID NO:
56);

and a corresponding or complementary DNA or RNA sequence,
wherein R=A or G, K=G or T/U, Y=C or T/U, W=A or T/U, N=A or C
or G or T/U, S=C or G, and B=C or G or T/U; and

-144-

the core region does not contain a stop codon in one or more reading frames.

44. The method of claim 33, wherein the first and/or second recombination site comprises at least a first nucleic acid sequence
5 selected from the group consisting of a mutated att recombination site containing at least one mutation that enhances recombinational specificity, a complementary DNA sequence thereto, and an RNA sequence corresponding thereto.

45. The method of claim 33, wherein the vector comprising the
10 second site further encodes at least one selectable marker.

46. The method of claim 45, wherein the marker is a promoterless marker, which, upon recombination is under the control of a promoter and is thereby expressed.

47. The method of claim 46, wherein the first recombination site
15 is attP and is in the sense orientation prior to recombination.

48. The method of claim 46, wherein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic, fluorescent, or capable of being bound by an antibody and FACs sorted.

20 49. The method of claim 48, wherein the selectable marker is selected from the group consisting of green fluorescent protein (GFP), red fluorescent protein (RFP), blue fluorescent protein (BFP), and *E. coli* histidinol dehydrogenase (hisD).

50. A cell comprising, the chromosome of claim 1.

25 51. The cell of claim 50, wherein the cell is a nuclear donor cell.

52. The cell of claim 50, wherein the cell is a stem cell.

53. The stem cell of claim 52, wherein said stem cell is human and is selected from the group consisting of a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell and an embryonic stem cell.

-145-

54. The cell of claim 50, wherein the cell is mammalian.

55. The cell of claim 54, wherein the mammal is selected from the group consisting of humans, primates, cattle, pigs, rabbits, goats, sheep, mice, rats, guinea pigs, hamsters, cats, dogs, and horses.

5 56. The cell of claim 50, wherein the cell is a plant cell.

57. A cell comprising the platform ACes of claim 14.

58. The cell of claim 57, wherein the cell is a nuclear donor cell.

59. The cell of claim 57, wherein the cell is a stem cell.

60. The stem cell of claim 59, wherein said stem cell is human
10 and is selected from the group consisting of a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell and an embryonic stem cell.

61. A human mesenchymal cell comprising an artificial chromosome.

62. The human mesenchymal cell of claim 61, wherein said
15 artificial chromosome is an ACes.

63. The human mesenchymal cell of claim 62, wherein the ACes is a platform-ACes.

64. A method for introducing heterologous nucleic acid into the mesenchymal cell of claim 63, comprising:

20 (a) introducing into the cell of claim 63, wherein the platform-ACes has a first recombination site, a vector comprising at least a second recombination site and the heterologous nucleic acid;

(b) incubating the resulting mixture in the presence of at least one recombination protein under conditions whereby recombination between
25 the first and second recombination sites is effected, thereby introducing the heterologous nucleic acid into the platform-ACes within the mesenchymal cell.

65. A lambda-intR mutein comprising a glutamic acid to arginine change at position 174 of wild-type lambda-intR.

-146-

66. The lambda-intR mutein of claim 65, wherein the lambda-intR mutein comprises SEQ ID NO:37.

67. The method of claim 46 wherein the promoterless marker is transcriptionally downstream of the heterologous nucleic acid, wherein
5 the heterologous nucleic acid encodes a heterologous protein, and wherein the expression level of the selectable marker is transcriptionally linked to the expression level of the heterologous protein.

68. The method of claim 67, wherein the selectable marker and the heterologous nucleic acid are transcriptionally linked by the presence
10 of a IRES between them.

69. The method of claim 68, wherein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic or fluorescent.

70. The method of claim 69, wherein the selectable marker is
15 selected from the group consisting of green fluorescent protein (GFP), red fluorescent protein (RFP), blue fluorescent protein (BFP), and *E. coli* histidinol dehydrogenase.

71. The method of claim 67 further comprising expressing the
20 heterologous protein and isolating the heterologous protein.

72. A method for producing a transgenic animal, comprising introducing a platform-ACes into an embryonic cell.

73. The method of claim 72, wherein the embryonic cell is a stem cell.

74. The method of claim 72, wherein the embryonic cell is in an
25 embryo.

75. The method of claim 72, wherein the platform-ACes comprises heterologous nucleic acid that encodes a therapeutic product.

-147-

76. The method of claim 72, wherein the transgenic animal is a fish, insect, reptile, amphibians, arachnid or mammal.

77. The method of claim 72, wherein the *ACes* is introduced by cell fusion, lipid-mediated transfection by a carrier system, microinjection,
5 microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer.

78. A transgenic animal produced by the method of claim 72.

79. A cell line useful for making a library of *ACes*, comprising a multiplicity of heterologous recombination sites randomly integrated
10 throughout the endogenous chromosomes.

80. A method of making a library of *ACes* comprising random portions of a genome, comprising introducing one or more *ACes* into the cell line of claim 79, under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of
15 the heterologous recombination sites within the cell's chromosomal DNA; and isolating said multiplicity of *ACes*, thereby producing a library of *ACes* whereby multiple *ACes* have different portions of the genome within.

81. A library of cells useful for genomic screening, said library
20 comprising a multiplicity of cells, wherein each cell comprises an *ACes* having a mutually exclusive portion of a chromosomal nucleic acid therein.

82. The library of cells of claim 81, wherein the cells of the library are from a different species than the chromosomal nucleic acid
25 within the *ACes*.

83. A method of making one or more cell lines, comprising
a) integrating into endogenous chromosomal DNA of a selected cell species, a multiplicity of heterologous recombination sites,

-148-

b) introducing a multiplicity of *ACes* under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites integrated within the cell's endogenous chromosomal DNA;

5 c) isolating said multiplicity of *ACes*, thereby producing a library of *ACes* whereby a multiplicity of *ACes* have mutually exclusive portions of the endogenous chromosomal DNA therein;

d) introducing the isolated multiplicity of *ACes* of step c) into a multiplicity of cells, thereby creating a library of cells;

10 e) selecting different cells having mutually exclusive *ACes* therein and clonally expanding or differentiating said different cells into clonal cell cultures, thereby creating one or more cell lines.

84. The method of claim 23, wherein the nucleic acid molecule with a recombination site is a PCR product.

15 85. Method of claim 23 wherein the recombinase is a protein and the recombination event occurs in vitro.

86. The method of claim 33, wherein the vector is a PCR product comprising a second recombination site.

20 87. The lambda-intR mutein of claim 65, wherein the mutein further comprises an amino acid signal for nuclear localization.

88. The lambda-intR mutein of claim 65, wherein the mutein further comprises an epitope tag for protein purification.

89. A modified iron-induced promoter comprising SEQ ID NO:128.

25 90. A plasmid or expression cassette comprising the promoter of claim 89.

91. A vector, comprising:
a recognition site for recombination; and

-149-

a sequence of nucleotides that targets the vector to an amplifiable region of a chromosome.

92. The vector of claim 91, wherein the amplifiable region comprises heterochromatic nucleic acid.

5 93. The vector of claim 91, wherein the amplifiable region comprises rDNA.

94. The vector of claim 93, wherein the rDNA comprises an intergenic spacer.

95. The vector of claim 91, further comprising nucleic acid
10 encoding a selectable marker that is not operably associated with any promoter.

96. The vector of claim 91, wherein the chromosome is a mammalian chromosome.

97. The vector of claim 91, wherein the chromosome is a plant
15 chromosome.

98. A cell of claim 57 that is a plant cell, wherein the ACes platform is a MAC.

99. The plant cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

20 100. The plant cell of claim 99, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

101. A cell of claim 57 that is an animal cell, wherein the ACes platform is a plant artificial chromosome (PAC).

25 102. The cell of claim 101 that is a mammalian cell.

103. The cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

104. The cell of claim 102, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

-150-

105. The cell of claim 104, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

106. A method, comprising:
5 introducing a vector of claim 91 into a cell;
growing the cells; and
selecting a cell comprising an artificial chromosome that comprises one or more repeat regions.

107. The method of claim 106, wherein sufficient portion of the
10 vector integrates into a chromosome in the cell to result in amplification of chromosomal DNA.

108. The method of claim 106, wherein the artificial chromosome is an *ACes*.

109. A method for screening, comprising:
15 contacting a cell comprising a reporter *ACes* with test compounds or known compounds, wherein:

the reporter *ACes* comprises one or a plurality of reporter constructs;

a reporter construct comprises a reporter gene in operative linkage
20 with a regulatory region responsive to test or known compounds; and
detecting any increase or decrease in signal output from the reporter, wherein a change in the signal is indicative of activity of the test or known compound on the regulatory region.

110. The method of claim 109, wherein the reporter is operatively
25 linked to a promoter that controls expression of a gene in a signal transduction pathway, whereby activation or reduction in the signal indicates that the pathway is activated or down-regulated by the test compound.

-151-

111. The method of claim 109, wherein the reporter in the construct encodes drug resistance or encodes a fluorescent protein.

112. The method of claim 111, wherein the fluorescent protein is selected from the group consisting of red, green and blue fluorescent
5 proteins.

113. The method of claim 109, wherein the ACes comprises a plurality of reporter-linked constructs, each with a different reporter, whereby the pathway(s) affected by the test compounds can be elucidated.

10 114. The method of claim 109, wherein a reporter is operatively linked to a promoter that is transcriptionally regulated in response to DNA damage, and the test compounds are genotoxicants.

115. The method of claim 114, wherein the DNA damage is induced by apoptosis, necrosis or cell-cycle perturbations.

15 116. The method of claim 114, wherein unknown compounds are screened to assess whether they are genotoxicants.

117. The method of claim 114, wherein the promoter is a cytochrome P450-profiled promoter.

20 118. The method of claim 114, wherein the cell is in a transgenic animal and toxicity is assessed in the animal.

119. The method of claim 109, wherein:
the cell is a patient cell sample; the patient has a disease;
the regulatory region is one targeted by a drug or drug regimen;
and

25 the method assesses the effectiveness of a treatment for the disease for the particular patient.

120. The method of claim 119, wherein the cell is a tumor cell.

121. The method of claim 109, wherein the cell is a stem cell or a progenitor cell, whereby expression of the reporter is operatively linked to

-152-

a regulatory region expressed in the cells to thereby identify stem cells or progenitor cell.

122. The method of claim 109, wherein the cell is in an animal;
and the method comprises whole-body imaging to monitor expression of
5 the reporter in the animal.

123. A reporter *ACes* comprises one or a plurality of reporter constructs, wherein the reporter construct comprises a reporter gene in operative linkage with a regulatory region responsive to test or known compounds.

Fig. 1 Generation of ACes for Platform
Chromosome Engineering

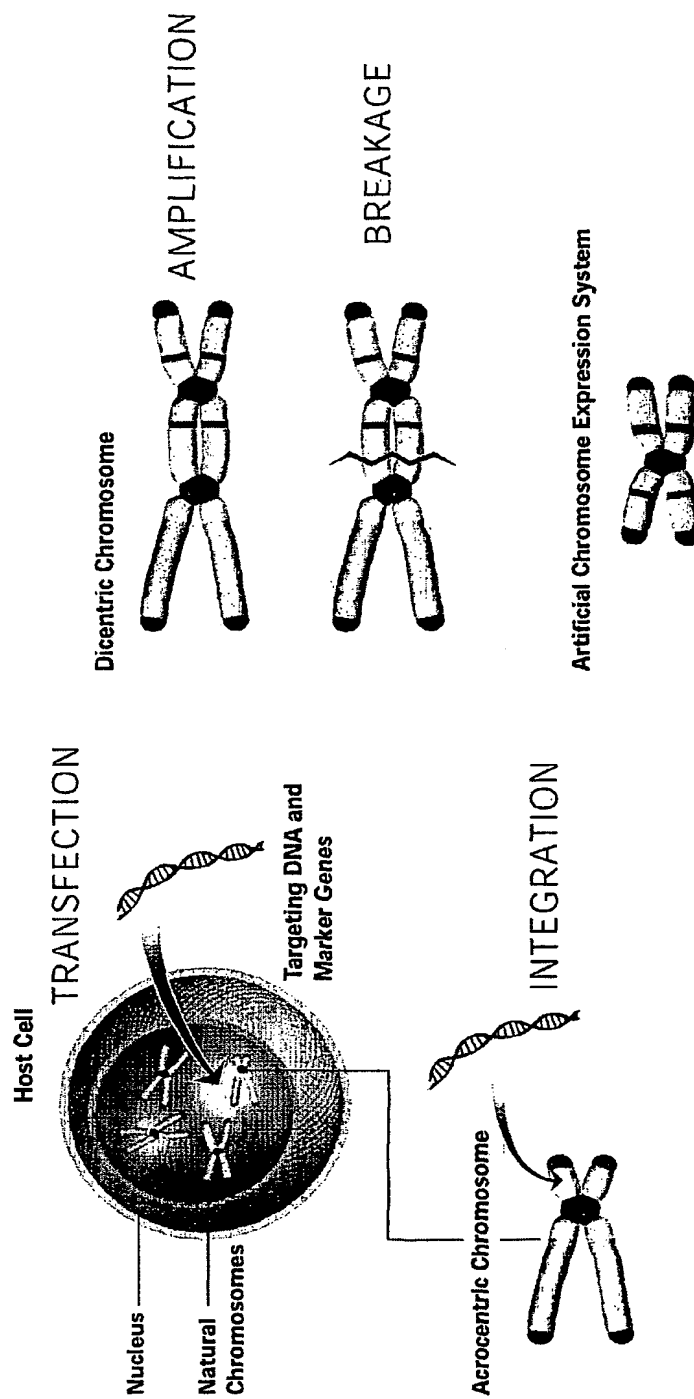
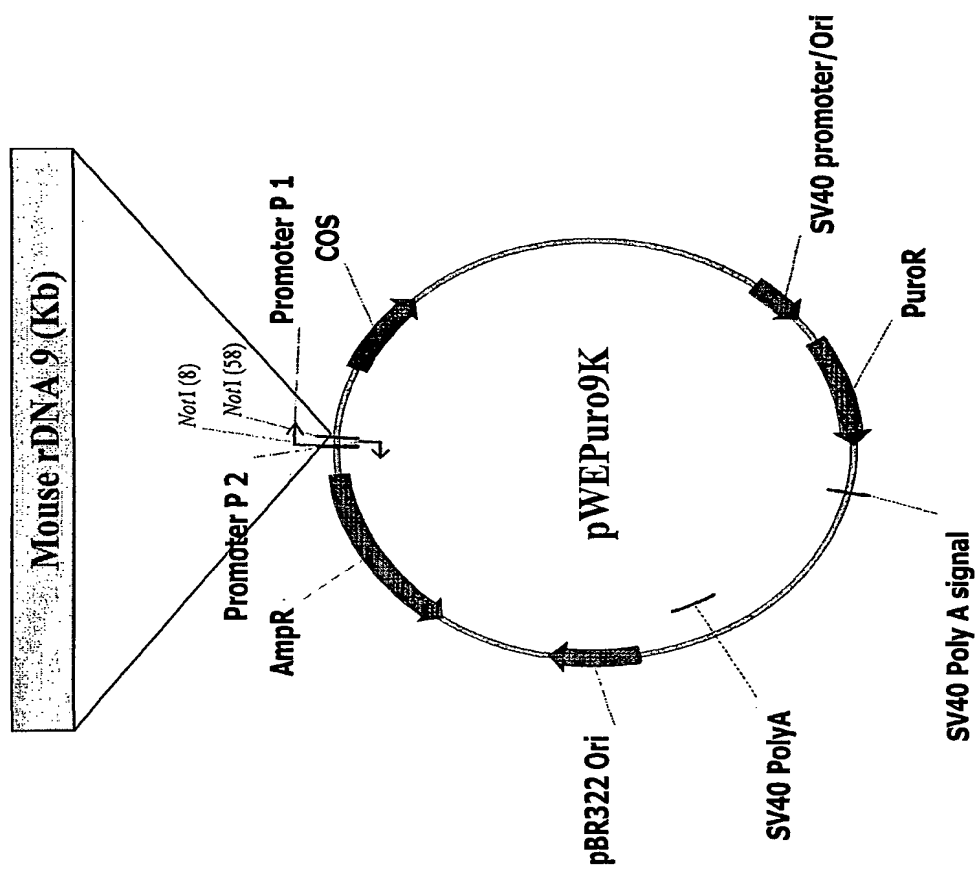


Fig. 2 Targeting plasmid pWEPuro9K



3/15

Fig. 3 Platform Chromosome with single recombination site

Platform chromosome containing mouse rDNA, mouse major/minor repeat DNA and puromycin resistant cassette



Targeting onto platform chromosome via homologous recombination



or



Engineered Selectable Marker System

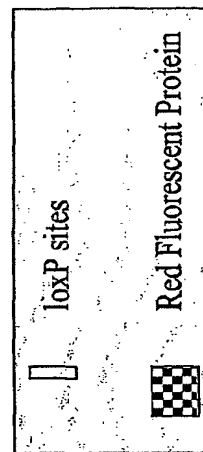
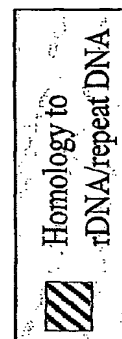
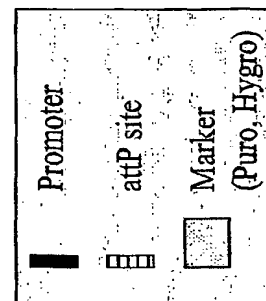
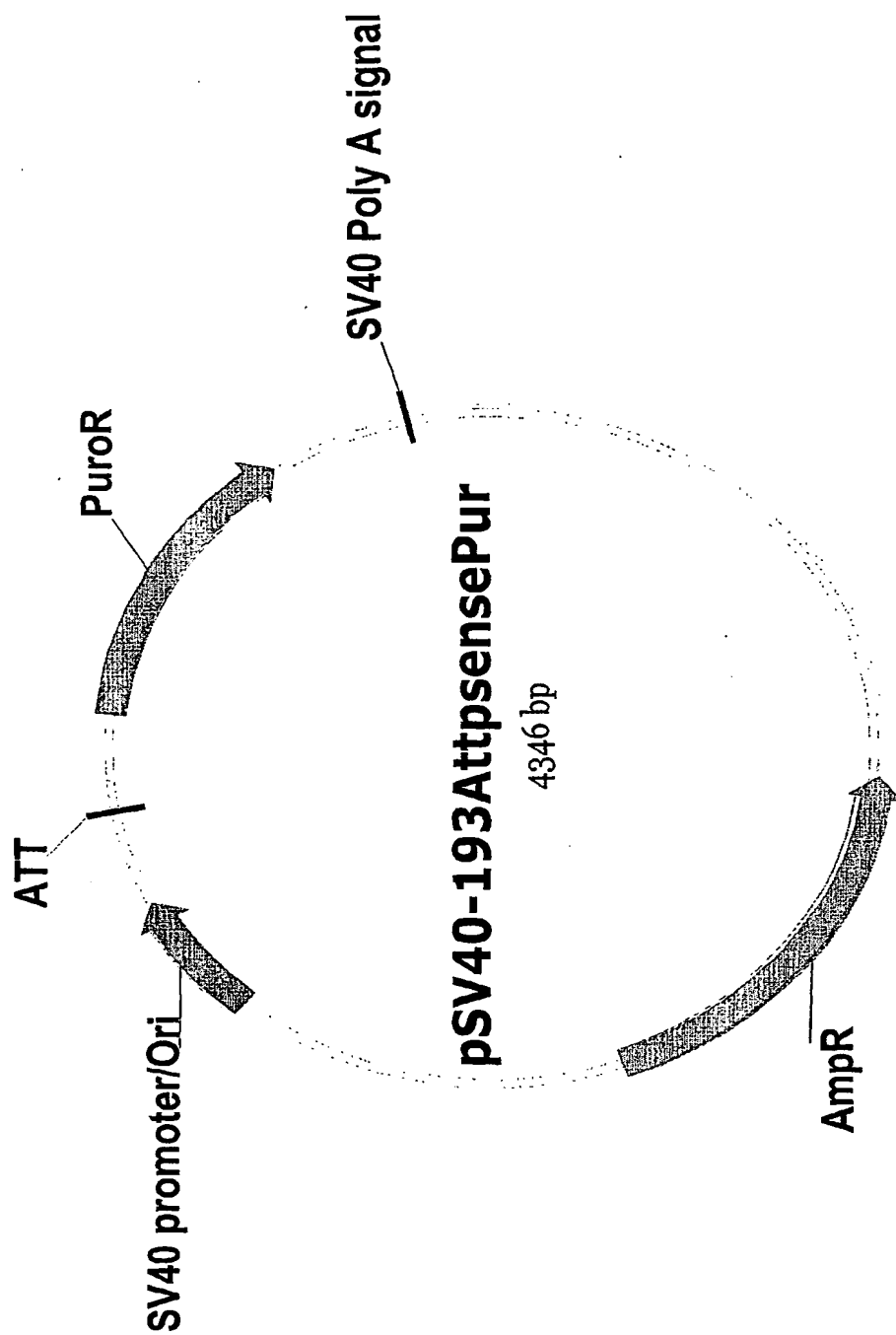


Fig. 4



5/15

Fig. 5 Generation of chromosome based platform with multiple integration sites

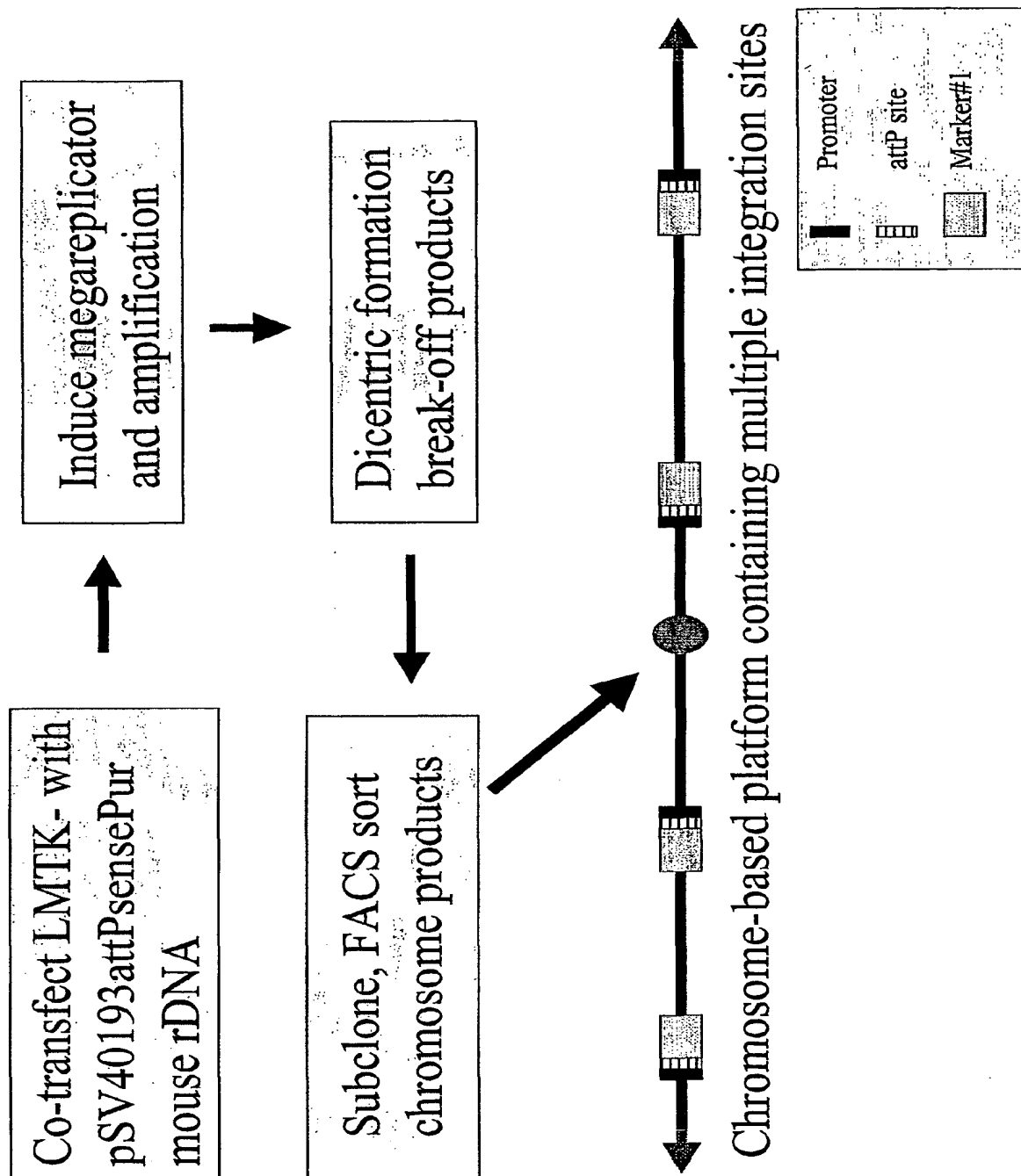


Fig. 6 λ integrase recombination

Core Region	
attP	CAGCTTTT <u>T</u> TATACTAA <u>G</u> TTG
attB	CTGCTTTT <u>T</u> TATACTAA <u>G</u> TTG
attL	CTGCTTTT <u>T</u> TATACTAA <u>G</u> TTG
attR	CAGCTTTT <u>T</u> TATACTAA <u>G</u> TTG

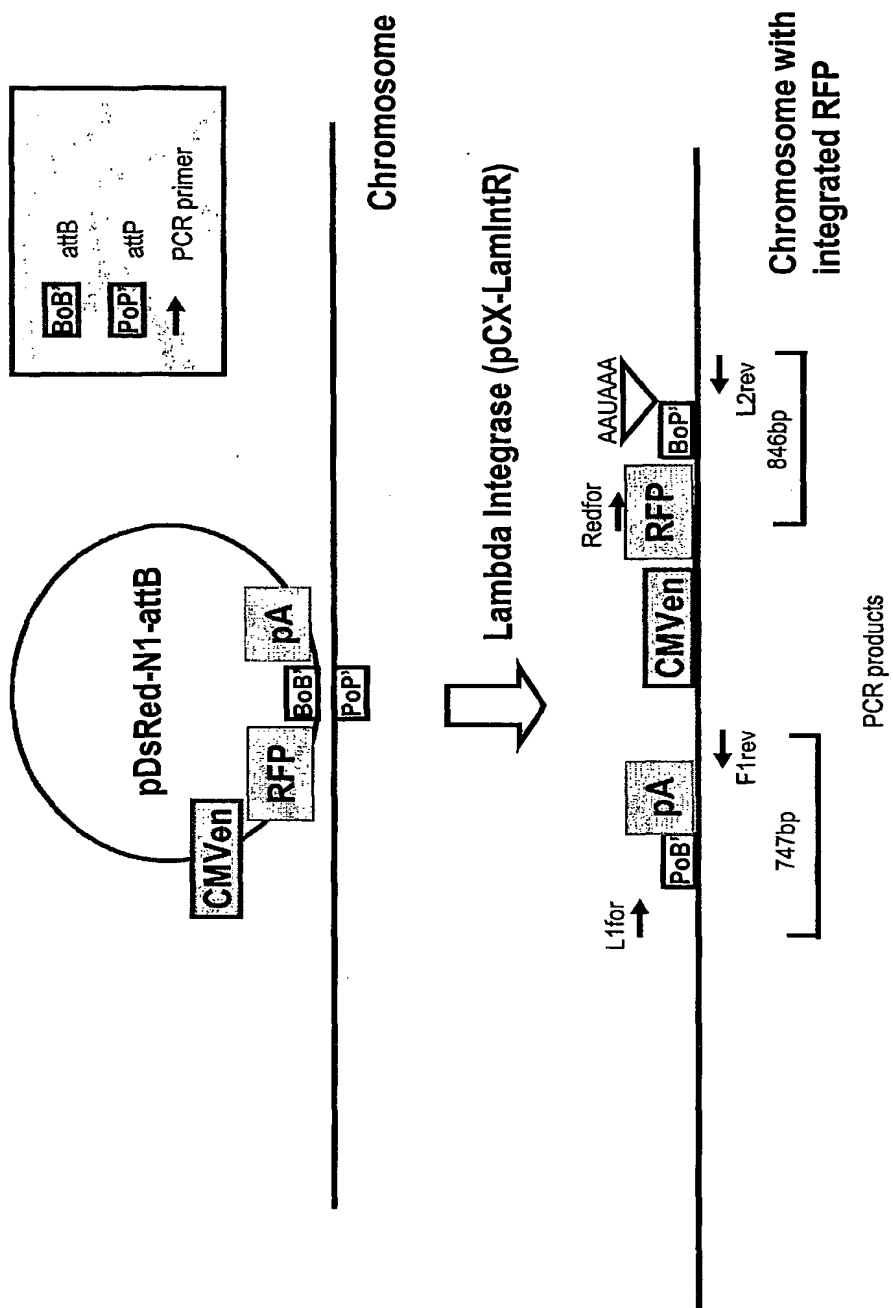
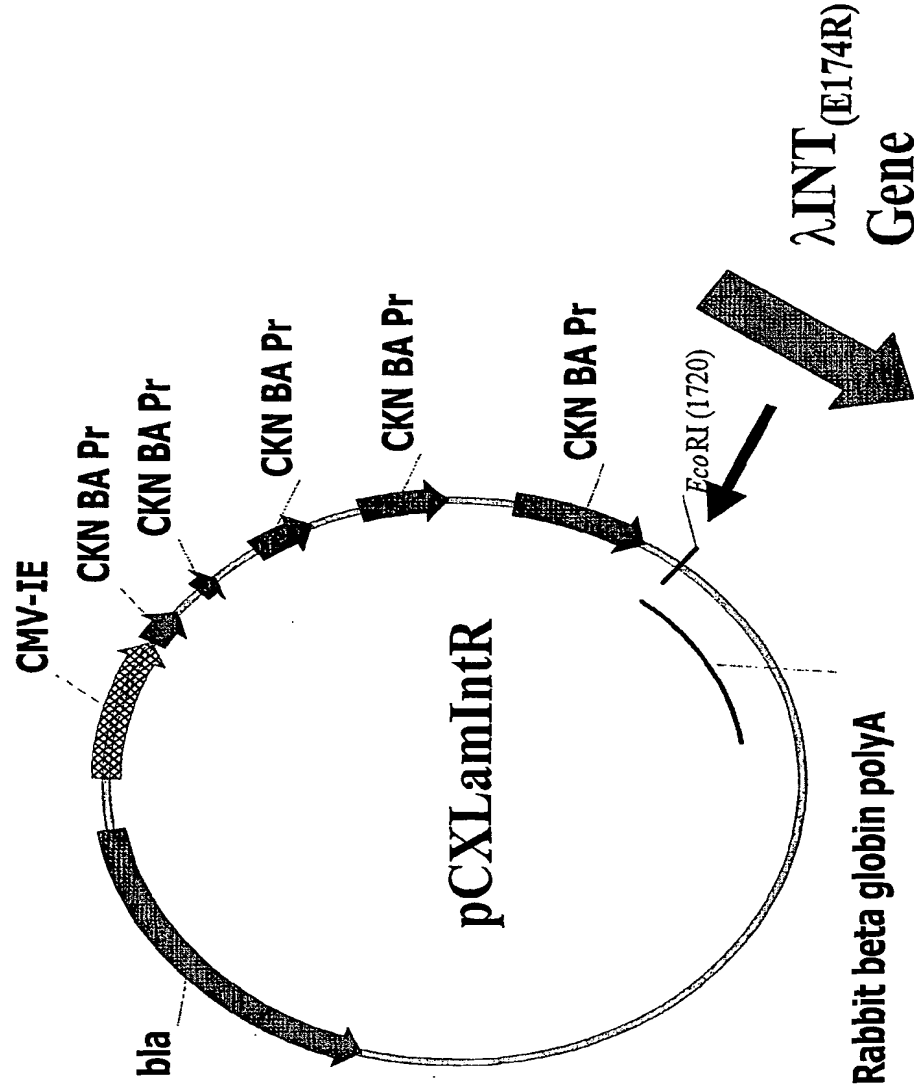
Fig. 7 λ INT recombination on artificial chromosome

Fig. 8 pCXLamIntR integrase expression vector



9/15

Fig. 9 Platform Technology

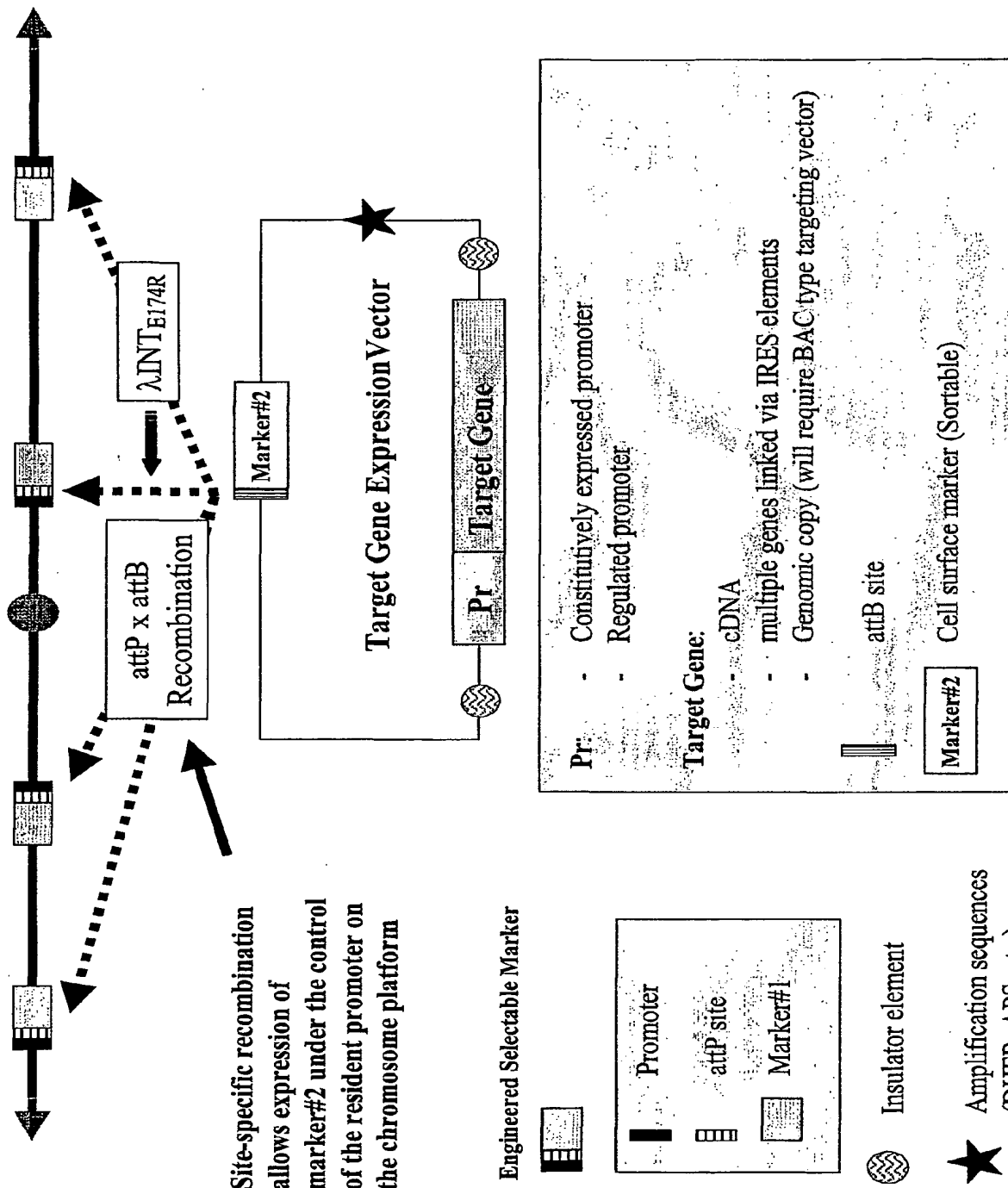


Fig. 10

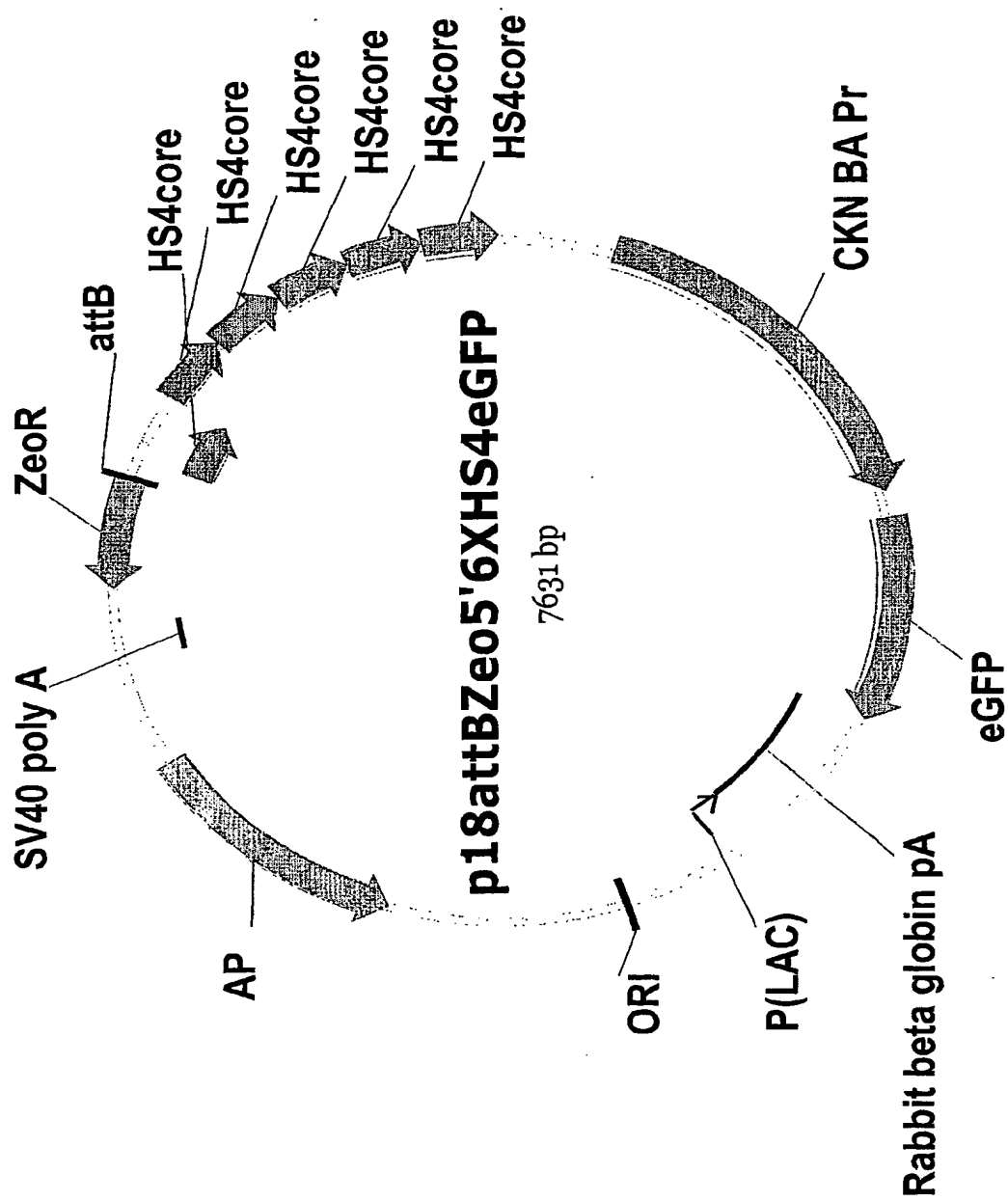


Fig. 11

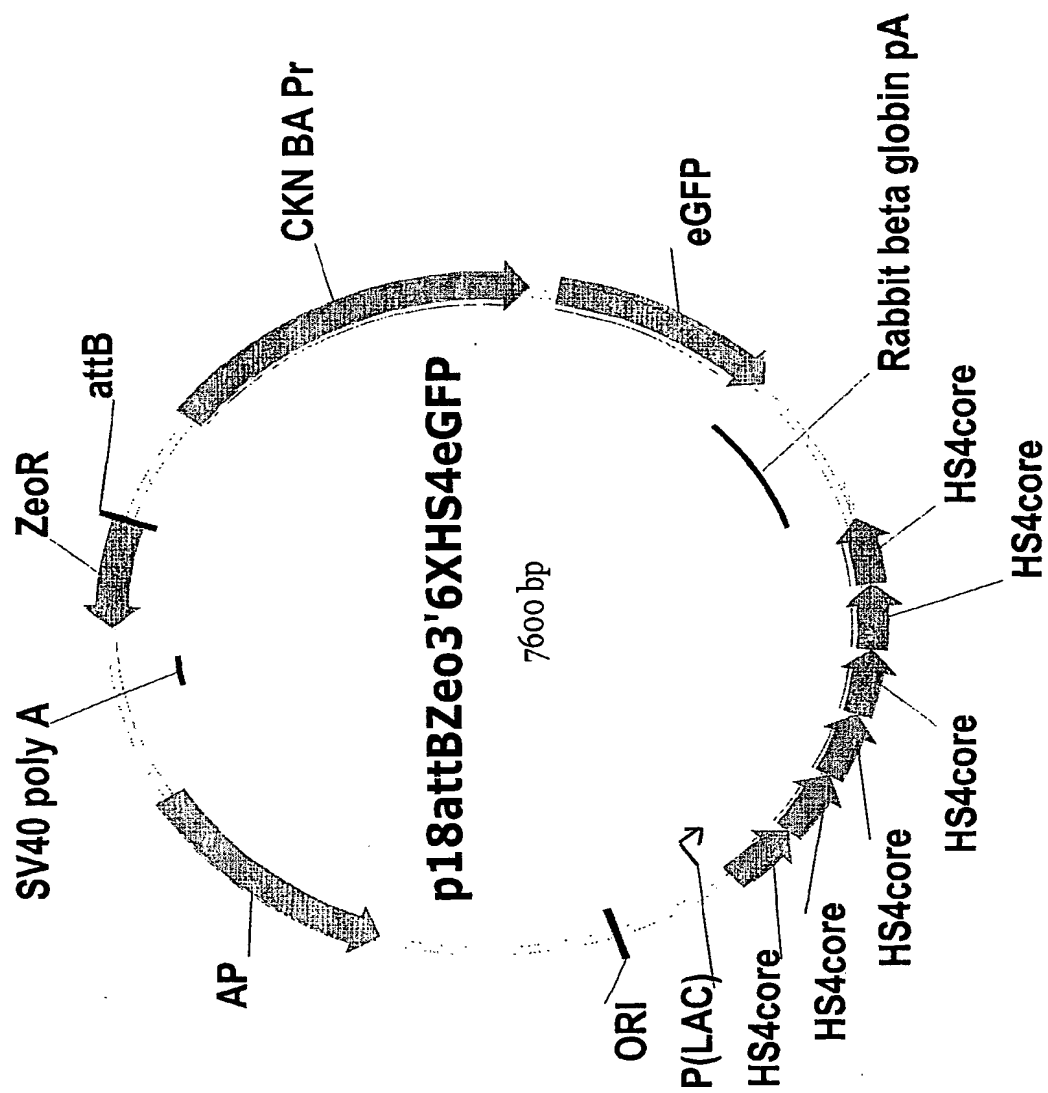


Fig. 12

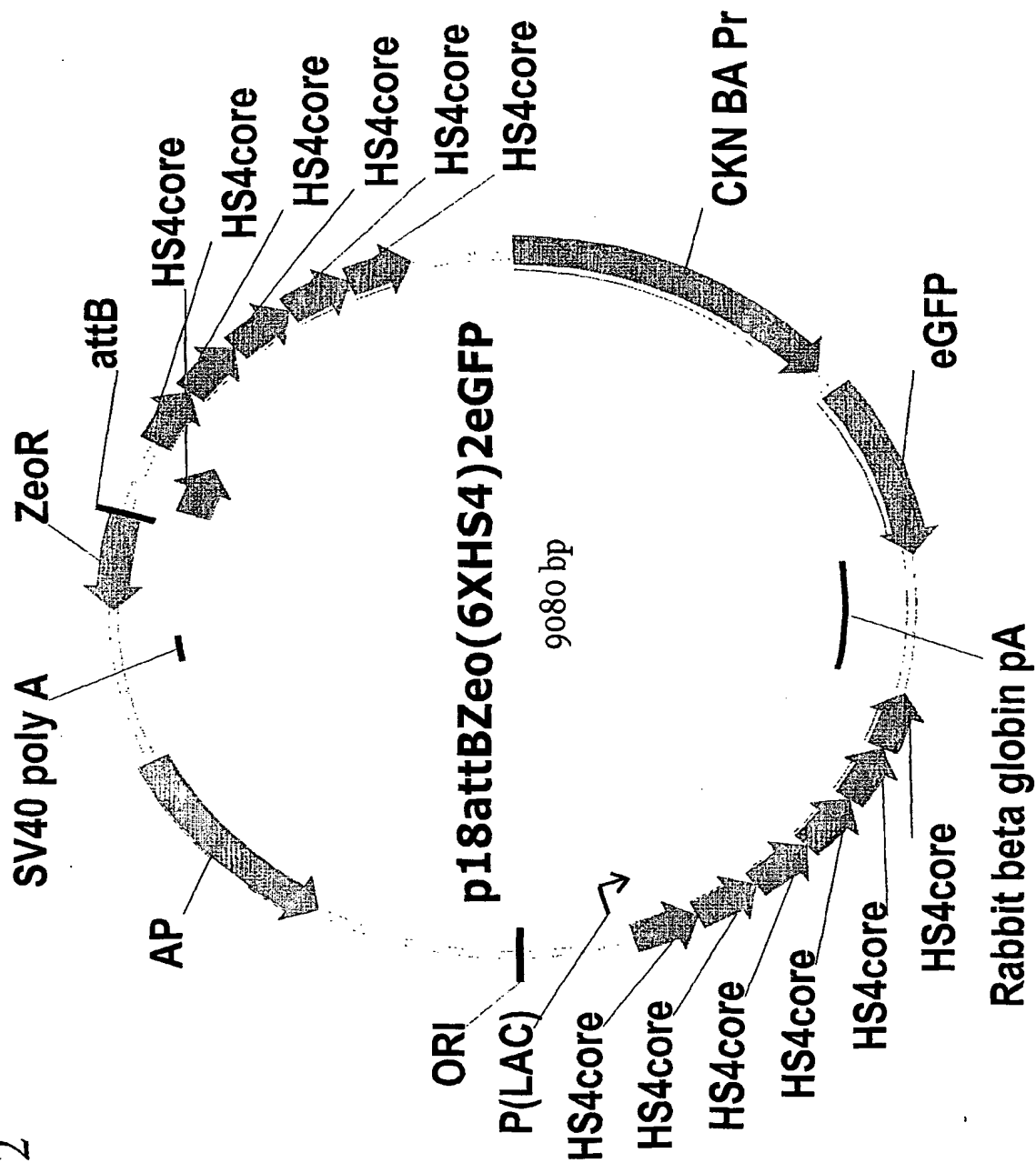
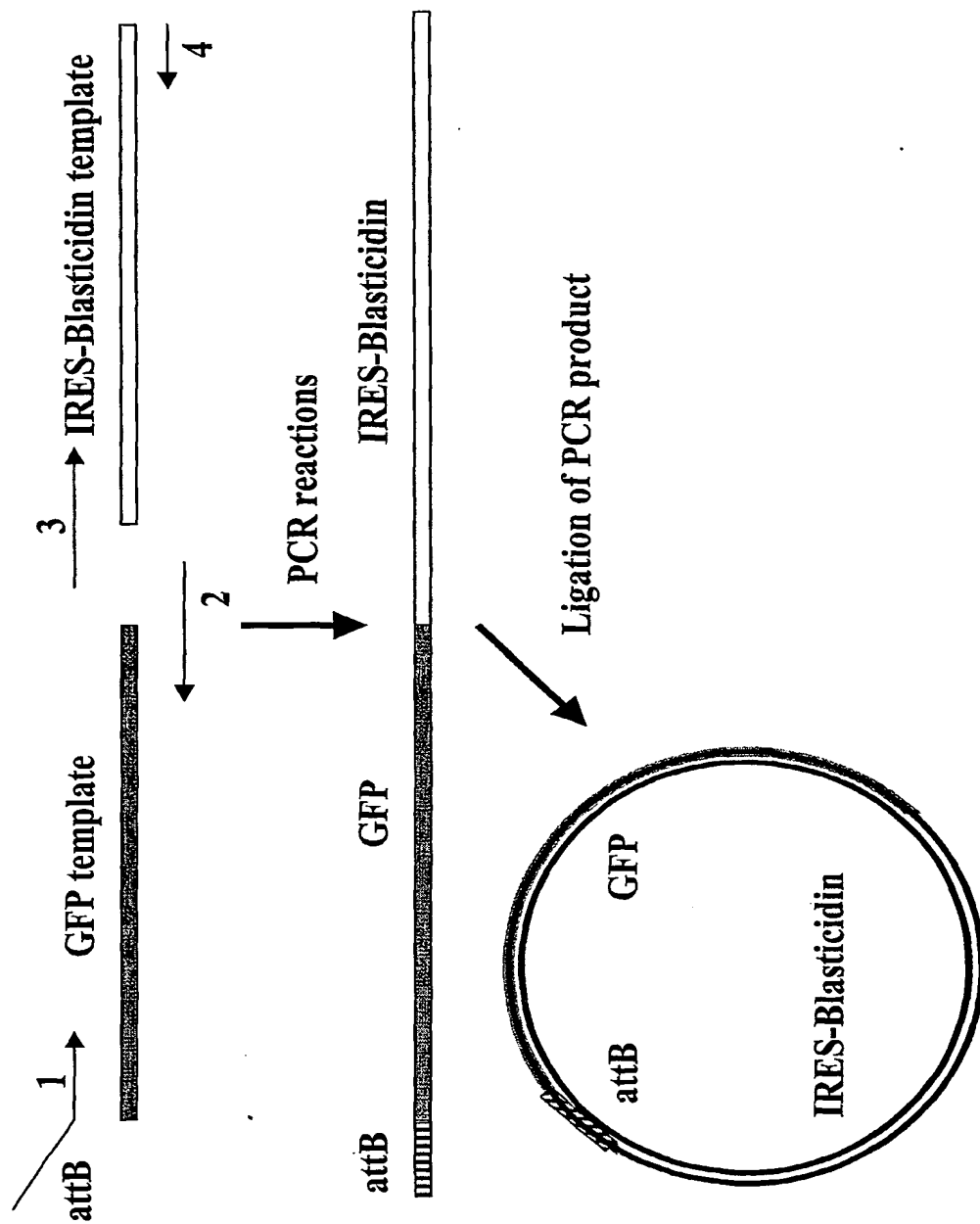


Fig. 13



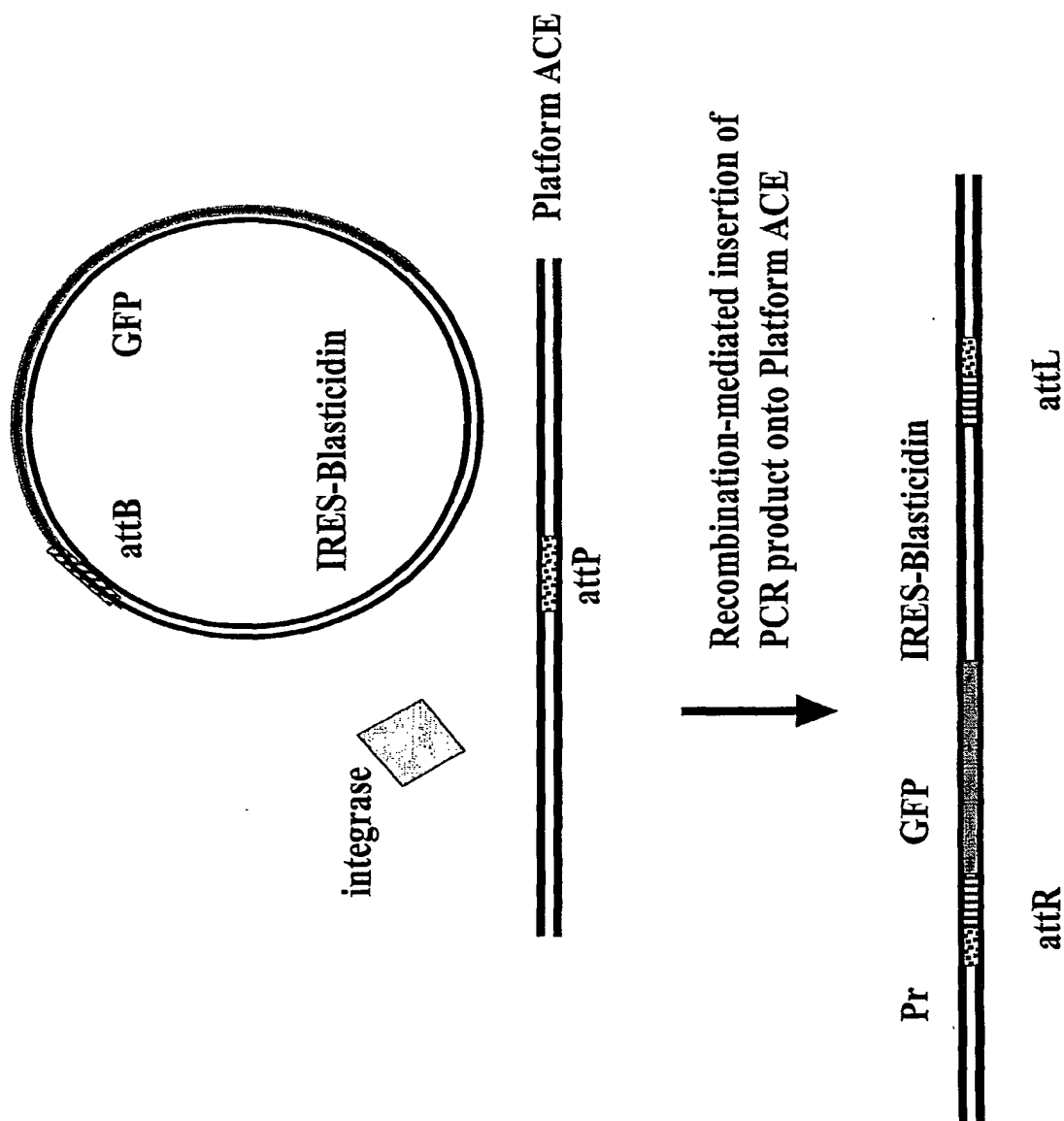
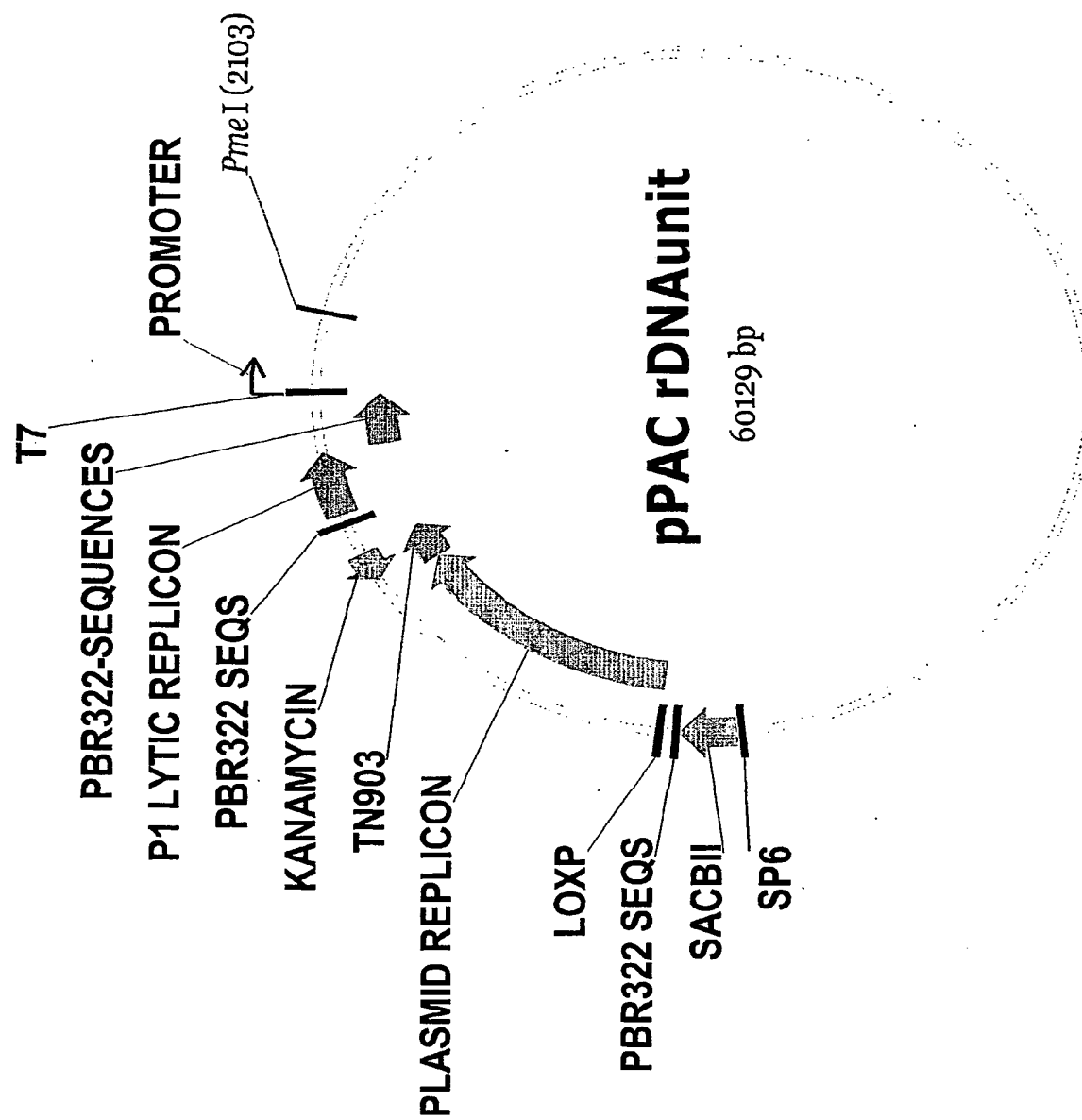


Fig. 14



-1-

SEQUENCE LISTING

<110> CHROMOS MOLECULAR SYSTEMS, INC.
Perkins, Edward
Perez, Carl
Lindenbaum, Michael
Greene, Amy
Leung, Josephine
Fleming, Elena
Stewart, Sandra
Shellard, Joan

<120> CHROMOSOME-BASED PLATFORMS

<130> 24601-420PC

<140> Not Yet Assigned
<141> Herewith

<150> 60/294,758
<151> 2001-05-30

<150> 60/366,891
<151> 2002-03-21

<160> 129

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer: attPUP

<400> 1
ccttgcgcta atgctctggt acagg 25

<210> 2
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer: attPDWN

<400> 2
cagaggcagg gagtgggaca aaattg 26

<210> 3
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer: Lamint 1

<400> 3
ttcgaattca tgggaagaag gcgaagtcac gagcg 35

<210> 4
<211> 34
<212> DNA
<213> Artificial Sequence

-2-

<220>
<223> Primer: Lamint 2

<400> 4
ttcgaattct tatttgattt caattttgtc ccac 34

<210> 5
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 5
cggacaatgc ggttgtgcgt 20

<210> 6
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 6
cgcgagcaa aatctagagt aaggagatca agacttacgg ctgacg 46

<210> 7
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> LambdaINTER174rev

<400> 7
cgtcagccgt aagtcttgat ctccttactc tagattttgc tgcgcg 46

<210> 8
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> attB1

<400> 8
tgaagcctgc ttttttatac taacttgagc gaa 33

<210> 9
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> attB2

<400> 9
ttcgctcaag ttagtataaa aaagcaggct tca 33

<210> 10
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

-3-

<223> Primer: attPdwn2

<400> 10
tcttctcggg cataagtcgg acacc 25

<210> 11
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:CMVen

<400> 11
ctcacgggga tttccaagtc tccac 25

<210> 12
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:attPdwn

<400> 12
cagaggcagg gagtgggaca aaattg 26

<210> 13
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:CMVEN2

<400> 13
caactccgcc ccattgacgc aaatg 25

<210> 14
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:L1

<400> 14
agtatcgccg aacgattagc ttttca 26

<210> 15
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:F1 rev

<400> 15
gccgatttcg gcctattggt taaa 24

<210> 16
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:RED

-4-

<400> 16
ccgccgacat ccccgactac aagaa 25

<210> 17
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:L2rev

<400> 17
ttccttcgaa ggggatccgc ctacc 25

<210> 18
<211> 22118
<212> DNA
<213> Mus musculus

<300>
<308> GenBank X82564
<309> 1996-04-09

<400> 18
gaattccctc atccctaata cagattgggtg gaataacttg gtatagatgt ttgtgcatta 60
aaaaccctgt aggatcttca ctctaggtca ctgttcagca ctggaacctg aatttgtggcc 120
ctgagtgata ggtcctggga catatgcagt tctgcacaga cagacagaca gacagacaga 180
cagacagaca gacagacgtt acaaacaaac acgttgagcc gtgtgccaac acacacacaa 240
acaccactct ggccataatt attgaggacg ttgatttatt attctgtgtt tgtgagtcctg 300
tctgtctgtc tgtctgtctg tctgtctgtc tatcaaacca aaagaaacca aacaattatg 360
cctgcctgcc tgcctgcctg cctacacaga gaaatgattt cttcaatcaa tctaaaacga 420
cctcctaagt ttgccttttt tctctttctt tatctttttc ttttttcttt tcttcttctt 480
tcttctcttc cttccttctt tcttctcttt ctttctttct ttctttcttt cttactttct 540
ttctttcttt cttacattta ttcttttcat acatagtttc ttagtgtaag catccctgac 600
tgtcttgaag acactttgta ggctcctaat ctgtaagagc cttcctctgc ttttcaaattg 660
ctggcatgaa tgttgtacct cactatgacc agcttagtct tcaagtctga gttactggaa 720
aggagtcca agaagactgg ttatattttt cattttatt tgcattttta ttaaaattta 780
atttcaccaa aagaatttag actgaccaat tcagagtctg ccgtttaaaa gcataaggaa 840
aaagtaggag aaaaacgtga ggctgtctgt ggatggctga ggctgcttta gggagcctcg 900
tcaccattct gcaacttgcaa accgggccac tagaaccogg tgaagggaga aaccaaagcg 960
acctggaac aataggtcac atgaaggcca gccacctcca tcttggttggt cgggagttca 1020
gttagcagac aagatggctg ccatgcacat gttgtctttc agcttggtga ggtcaaagta 1080
caaccgagtc acagaacaag gaagtataca cagtgtgttc caggtcagcc agagtttaca 1140
cagagaaacc acatcttgaa aaaaacaaaa aaataaatta aataaatata atttaaaat 1200
ttaaaaatag ccgggagtgaa tggcgcatgt ctttaatccc agctctcttc aggcagagat 1260
gggaggattt ctgagtttga ggccagcctg gtctgcaaag tgagttccag gacagtcagg 1320
gctatacaga gaaacctgt cttgaaaact aaactaaatt aaactaaact aaactaaaaa 1380
aatataaaat aaaaatttta aagaatttta aaaaactaca gaaatcaaac ataagcccac 1440
gagatggcaa gtaactgcaa tcatagcaga aatattatac acacacacac acacagactc 1500
tgtcataaaa tccaatgtgc cttcatgatg atcaaatttc gatagtcagt aatactagaa 1560
gaatcatatg tctgaaaata aaagccagaa ccttttctgc ttttggtttc ttttgcccca 1620
agatagggtt tctctcagtg tatccttggc atcctgcctt ggaacttctt ttgtaggttt 1680
ggtagcctca aactcagaga ggtcctctct gcctgcctgc ctgcctgcct gcctgcctgc 1740
ctgcctgcct gcctgcctca cttcttctgc caccacaca accgagtcga acctaggatc 1800
tttatttctt tctctttctc tcttctttct ttctttcttt ctttctttct ttctttcttt 1860
ctttctttct ttcttattca attagttttc aatgtaagtg tgtgtttgtg ctctatctgc 1920
tgcctatagg cctgcttgcc aggagagggc aacagaacct aggagaaacc acctgacgc 1980
tctctgagaat aagtgaaaaa acaacaaaaa aggaaaattc taatcacata gaatgttagt 2040
atatgccgag gctgtcagag tgctttttta ggcttagtgt aagtaatgaa aatttgttgt 2100
tgtcttttat ccaaacacag aagagaggtg gctcggcctg catgtctgtt gtctgcagtt 2160
agaccaggct ggcttgaac acattaatct gctcgcctct gcttccctaa tgcctgcatt 2220
aaaggcatgt gccaccactg cccggactga tttcttcttt tttttttttt tggaaaatac 2280
ctttctttct ttttctctct ctctttctct cttcttctct ttctttctat tctttttttc 2340
tttctttttt cttttttttt ttttttttaa aatttgcta aggttaaggg tgtgctccac 2400
aattgcctca gctctgctct aattctcttt aaaaaaaaac aaacaaaaaa aaaacccaaa 2460
cagtatgtat gtatgtatat ttagaagaaa tactaatcca ttaataactc ttttttctta 2520
aaattcatgt cattcttggt ccacaaagtg agttccagga cttaccagag aaaccctgtg 2580

ttcaaatttc	tgtgttcaag	gtcaccctgg	cttacaaagt	gagttccaag	tccgataggg	2640
ctacacagaa	aaaccatac	tcagaaaaaa	aaaaagttcc	aaacacacac	acacacacac	2700
acacacacac	acacacacac	acacacacac	acacacacag	cgcgcgcggg	cgatgagggg	2760
aagtcgtgcc	taaaaataat	atttttctgg	cctaaagtga	agcaaatcac	tatgaagagg	2820
tactcctaga	aaaaataaat	acaaacgggc	tttttaatac	ttccagcact	gttttaattt	2880
aactctgaat	ttagtcttgg	aaaagggggc	gggtgtgggt	gagtgagggc	gagcgagcag	2940
acgggcgggc	gggcgggtga	gtggccggcg	gcggtggcag	cgagcaccag	aaaacaacaa	3000
accccaagcg	gtagagtgtt	ttaaaaatga	gacctaaatg	tgggtggaacg	gaggtcgccg	3060
ccaccctcct	ttccactgc	ttagatgctc	ccttcccctt	actgtgctcc	cttcccctaa	3120
ctgtgcctaa	ctgtgcctgt	tccctcacc	cgctgattcg	ccagcgacgt	actttgactt	3180
caagaacgat	tttgccctgt	ttcaccgctc	cctgtcatac	tttcgttttt	gggtgccccg	3240
gtctagcccg	ttcgctatgt	tcgggcggga	cgatggggac	cgtttgtgcc	actcgggaga	3300
agtgggtgggt	gggtacgctg	ctccgtcgtg	cgtgctgag	tgccggaacc	tgagctcggg	3360
agaccctccg	gagagacaga	atgagttagt	gaatgtggcg	gcgcgtgacg	gatctgtatt	3420
ggtttctgat	ttgatcagag	accattgtcg	ggcgacacct	agtggtgaca	agtttcggga	3480
acgctccagg	cctctcaggt	tggtgacaca	ggagagggaa	gtgcctgtgg	tgaggcgacc	3540
agggtgacag	gagggccggc	aagcagggcg	gagcgtctcg	gagatgggtg	cgtgtttaag	3600
gacggtctct	aacaaggagg	tcgtacaggg	agatggcaca	agcagaccga	gttgctgtac	3660
gcccttttgg	gaaaaatgct	aggggtgggt	gcaacgttac	taggtcgacc	agaaggctta	3720
agtcctaccc	ccccccccct	tttttttttt	tttctctcag	aagccctctc	ttgtccccgt	3780
caccgggggc	cccgctacac	tgaggccag	aggagccgat	gggcccggct	tccaagccgg	3840
tgtgggtcgg	ccagctggcg	cttcgggtct	tttttttttt	tttttttttt	ttttctctca	3900
gaagccctgt	ctgtcgctgt	caccgggggc	gctgtacttc	tgaggccgag	aggacgcgat	3960
gggcccccg	ttccaagccg	gtgtggctcg	gccagctgga	gcttcgggtc	tttttttttt	4020
tttttttttt	ttttttttct	cagaagcctt	gtctgtcgct	gtcaccgggg	gcgctgtact	4080
tctgaggccg	agaggacgcg	atgggtcgcc	ttccaagccg	atgtggccgg	gccagctgga	4140
gcttcgggtt	tttttttttt	ctccagaagc	cctctcttgg	ccccgtcacc	ggggcgctga	4200
tacttctgag	gccgagagga	cgatgatggg	ccgggttcca	ggcggatgtc	gcccggctcag	4260
ctggagcttt	ggatcttttt	tttttttttt	cctccagaag	ccctctcttg	tccccgtcac	4320
ggggggcacc	ttacatctga	gggcgagagg	acgtgatagg	tccggcttcc	aagccgatgt	4380
gcccggggcca	ctggagctt	cggttttttt	tttttctctc	cagaagccct	ctcttgtccc	4440
cgtcaccggg	ggcgctgtac	ttctgaggcc	gagaggacgt	gatgggcccc	ggttccaggc	4500
ggatgtcgcc	cggtcagctg	gagctttgga	tcattttttt	ttttccctcc	agaagccctc	4560
tcttgtcccc	gtcaccgggg	gcaccgtaca	tctgaggccg	agaggacacg	atgggcccgt	4620
cttccaagcc	gatgtggccc	ggccagctgg	agcttcgggt	cttttttttt	tttttccctc	4680
cagaagcctt	gtctgtcgct	gtcaccgggg	gcgctgtact	tctgaggccg	agaggacgcg	4740
atggggccgg	cttccaagcc	gggtgtggct	ggccagctgg	agcttcgggt	cttttttttt	4800
tttttttttt	ttcctccaga	aaccttctct	gtcgctgtca	ccccggggcg	ttgtacttct	4860
gatggccgag	ggacgcgatg	ggcccgctct	ccaggccgat	gtggcccggg	cagctggagc	4920
tttggatctt	tttttttttt	tttctctcca	gaagccctct	cttgtccccg	tcaccggggg	4980
caccttacat	ctgaggccct	gaggacacga	tgggcccggg	ttccaggccg	atgtggcccc	5040
gtcagctgga	gctttggatc	tttttttttt	ttttcttcca	gaagccctct	tgtccccctc	5100
accggtggca	ctgtacatct	gaggcggaga	ggacattatg	ggcccgggct	ccaatccgat	5160
gtggcccggg	cagctggagc	tttggatctt	attttttttt	taattttttt	ttccagaagc	5220
cctcttgtcc	ctgtcaccgg	tggcacggta	catctgaggg	cgagaggaca	ttatgggccc	5280
ggcttccagg	ccgatgtggc	ccggtcagct	ggagctttgg	atcttttttt	ttttttttct	5340
tttttccctc	agaagccctc	tctgtccctg	tcaccggggg	ccctgtacgt	ctgaggccga	5400
gggaaagcta	tgggcgcggg	tttctttcat	tgacctgtcg	gtcttatcag	ttctccgggt	5460
tgtcaggggtc	gaccagttgt	tcctttgagg	tcgggttctt	ttcgttatgg	ggtcattttt	5520
gggcccacctc	cccaggtagt	acttccaggc	gtcgttgctc	gcctgtcact	ttcctccctg	5580
tctcttttat	gcttgtgatc	ttttctatct	gttctatttg	gacctggaga	taggtactga	5640
cacgctgtcc	tttccctatt	aacactaaag	gacactataa	agagaccctt	tcgattttaag	5700
gctgttttgc	ttgtccagcc	tattcttttt	actggcttgg	gtctgtcgcg	gtgcctgaag	5760
ctgtccccga	gccacgcttc	ctgctttccc	gggcttgctg	cttgctgtgtg	cttgcgtgtg	5820
cgagcttgtg	acaactgggc	gctgtgactt	tgctgcgtgt	cagacgtttt	tcccgatctt	5880
cccaggtgtg	cgttgtcaca	cctgtcccgg	ttggaatggg	ggagccagct	gtgggtgagg	5940
gccaccttat	ttcggtcac	tttttttttt	tttttttctc	ttggagctcc	gaacctccgc	6000
tcttttctct	ttccgggtctt	tttccacat	gcctcccag	tgcatttctt	ttgttttttt	6060
ttcttttttt	tttttttttt	ttggggaggt	ggagagctcc	gagtaactca	ctcctgtctg	6120
tgggtgtccaa	gtgttcatgc	cacgtgcctc	ccgagtgac	ttttttttgt	ggcagtcgct	6180
cggtgtgttc	tgtctgccc	tgtctgccc	tatcagtaac	tgcttggccc	cgcgtgtaag	6240
acattcctat	ctcgttgtt	tctcccgatt	gcgcgtcggt	gctcactctt	agatcgatgt	6300
ggtgtctccg	agttctcttc	gggcccgggc	caagccgcgc	caggccaggg	acggacattc	6360
atggcgaatg	cgcccgctc	ttctcgttct	gccagcgggc	cctcgctctc	ccaccacatc	6420
cgtctgccc	tgggtgtgtg	aaggcagggg	tgcggctctc	cgcccccagc	ctgcccccg	6480
cgcacttttc	tcagtgggtc	gcgtggctct	tgtggatgtg	tgaggccccc	gggtgtgccc	6540
tcacgtgttt	cactttgggtc	gtgtctcgct	tgaccatggt	cccagagtcg	gtggatgtgg	6600

ccgggtggcgt	tgcataccct	tcccgtctgg	tgtgtgcacg	cgctgtttct	tgtgaagcgtc	6660
gaggtgctcc	tggagcgttc	caggttttgtc	tcctaggtgc	ctgcttctga	gctgggtggg	6720
gcgctcccca	ttccctgggtg	tgccctccggt	gctccgtctg	gctgtgtgcc	ttcccgtttg	6780
tgtctgagaa	gcccgtgaga	gggggggtcga	ggagagaagg	aggggcaaga	cccccttct	6840
tcgtcgggtg	aggcgccac	cccgcgacta	gtacgcctgt	gcgtagggtc	ggtgctgagc	6900
ggtcgcggct	ggggttggaa	agttttctcga	gagactcatt	gctttcccg	ggggagcttt	6960
gagaggcctg	gctttcgggg	gggaccgggt	gcagggtctc	ccctgtccgc	ggatgctcag	7020
aatgcccttg	gaagagaacc	ttcctgttgc	cgcagacccc	cccgcgcggt	cgcccccggtg	7080
ttggcttctt	ggtttccctg	tgtgtctgtc	gcatgcatcc	tctctcggtg	gcccgggctc	7140
gtcgggggtt	tgggtccgtc	ccgcccctcag	tgagaaagtt	tccttctcta	gctatcttcc	7200
ggaaaggggtg	cgggcttctt	acgggtctcga	gggggtctctc	ccgaatgggtc	ccctggagggtg	7260
ctcgcacctc	gaaacgcctcc	cgcgcgcgca	gcgtttgtctc	tctcgtctac	cgcggcccgc	7320
ggcctccccc	ctccgagttc	ggggagggtat	cacgcggggc	agagcctgtc	tgtcgtcctg	7380
ccgttgcctg	ggagcatgtg	gctcggcttg	tgtggttggg	ggctggggag	agggctccgt	7440
gcacaccccc	gcgtgcgcgt	actttctctc	cctcctgagg	gccgcggtgc	ggacggggctc	7500
tgggtaggcg	acgggtgggtc	cccggttccc	caccgctctt	cccgtgcctc	acccgtgcct	7560
tccgtcgcgt	gcgtccctct	cgctcgcgtc	cacgactttg	gccgctcccg	cgacggcggc	7620
ctgcgcgcgc	gctgtgtgct	gctgtgtgct	cctcgggctg	tgtggttgtg	tcgctcgc	7680
cccccttctc	cgcggcagcg	ttcccaaggc	tggcgaatc	gcgggagtc	tccttcccct	7740
cctcgggggtc	gagaggggtcc	gtgtctggcg	ttgattgatc	tcgctctcgg	ggacggggacc	7800
gttctgtggg	agaacggctg	ttggccgctg	cgcgcgcgac	gtcggacgtg	gggaccactc	7860
gccgctcggg	ggtcttcgtc	ggttaggcac	ggtgtgtcgg	catcggtctc	tctctcgtgt	7920
cggtgtcgcc	tcctcgggtc	cccggggggc	cgctgtgttt	cggtcgggt	cggcgtgca	7980
ggtgtgggtg	gactgctcag	gggagtgggtg	cagtgtgatt	cccgcgggtt	ttgcctcgcg	8040
tgccttgacc	ggtccgacgc	ccgagcgggtc	tctcgggtccc	ttgtgaggac	ccccctccgg	8100
gaggggcccc	tttcggcgcg	ccttgcgcgtc	gtcgcgggct	ctcgttctgc	tgtgtcgttc	8160
ccccctcccc	gctcgcgca	gcccgtcttt	tttctctct	ccccccctct	cctctgactg	8220
accggtggcc	gtgctgtcgg	accccccgca	tgggggcggc	cgggcacgta	cgcgctccggg	8280
cggtcacccg	ggtcttgggg	gggggcgcgag	gggtaagaaa	gtcgggtcgg	cgggcggggg	8340
gagctgtggt	ttggaggcg	tcccggcccc	cgcgcgcgtg	cggtgtcttg	cgcggtcttg	8400
gagaggggtg	cgtgcgaggg	gaaaagggtg	ccccgcgagg	gcaaagggaa	agaggttagc	8460
agtggctatt	gtcccgcagg	tgtggtgggtc	tgttggccga	ggtgcgtctg	gggggtcgt	8520
ccggccctgt	gaaggcgcgt	gttggggcct	gcccgggccc	gcccgaagtgc	cgaggtgggtg	8580
accctggcgg	tgggattaac	cccgcgcgcg	tgtcccgggtg	tggcgggtggg	ggctccgggtc	8640
ctcgtctacc	tcctctctcc	cgaggtctca	ggccttctcc	gcgcggggtc	tcggccctcc	8700
cctcgttccc	ccctctcgcg	gggttcaagt	cgtcgtcga	cctccccctc	tcgctcctcc	8760
catctctcgc	gcaatggcgc	cgcccagatt	cacgggtgggt	tcgtcctccg	cctccgcttc	8820
tcgcgggggg	ctggccgctg	tccggtctct	cctgcccagc	ccccgttggc	gtggtcttct	8880
ctcgcgggtc	ctggcctcgc	cggggttcgc	cggagggttc	agggggcttc	ccggttcccc	8940
gacgttgcgc	ctcgtgctg	tgtgcttggg	gggggcccgc	tgcggccctc	gcccgcgggt	9000
gagccccctg	cgcacccgcc	ggtgtgcggg	ttcgcgcgcg	ggtcagttgg	gcccgtggcg	9060
tgtgtcgcgt	cgggagcgtg	tccgcctcgc	ggcgggtaga	cgcgggtgtc	gcccgggtctc	9120
gacgggtggc	ctatccaggg	ctcgcgcccg	cgcacccccg	cctgcccgtc	ccggtgggtg	9180
tcgttgggtg	ggggagtga	tgggtgctacc	ggtcattccc	tcccgcgtgg	tttgactgtc	9240
tcgcgggtgt	cgcgcttctc	tttccgccaa	ccccacgcgc	aaccaccac	cctgctctcc	9300
cggcccggtg	cggtcgacgt	tccggtctct	ccgatgcgga	ggggttcggg	atttgtgccc	9360
gggacggagg	ggagagcggg	taagagaggt	gtcggagagc	tgtcccgggg	cgacgctcgg	9420
ggtggctttg	ccgctgcgt	gtgctcgcgg	acgggttttg	tcggaccccg	acgggggtcgg	9480
tcggcccgca	tgcactctcc	cggtccgcgc	gagcgcgcgc	ccgggtcacc	cccggtttgt	9540
cctcccgca	ggtctctcgc	cgcgcgcgc	tcctcctcct	ctctcgcgt	ctctgtcccg	9600
cctggtcctg	tcccaccccc	gacgctccgc	tcgcgcttcc	ttacctgggt	gacctgcca	9660
ggtagcatat	gcttgtctca	aagattaagc	catgcatgtc	taagtacgca	cggccgggtac	9720
agtgaactg	cgaatggctc	attaaatcag	ttatggttcc	tttgggtcgt	cgctcctctc	9780
ctacttggat	aactgtggta	attctagagc	taatacatgc	cgacggggcg	tgacccccct	9840
tcccgggggg	ggatgcgtgc	atttatcaga	tcaaaaccac	cccgttgagc	tccctcccg	9900
ctccggccgg	gggtcgggcg	ccggcgggtt	ggtgactcta	gataacctcg	ggccgatcgc	9960
acgccccccc	tggcggcgac	gacccattcg	aacgtctgcc	ctatcaactt	tcgatggtag	10020
tcgcccgtgc	taccattggtg	accacgggtg	ggatccattc	agggttcgat	tccggagagg	10080
gagcctgaga	aacggctacc	acatccaagg	aaggcagcag	gcgcgcaaat	taccactcc	10140
ggacccgggg	aggtagtgac	gaaaaataac	aatacaggac	tctttcgagg	ccctgtaatt	10200
cgaattgagtc	cctttaacga	cctttacattg	cttagctgag	gagggcaagt	ccctggccag	10260
cagccgcgggt	aattccagct	ccaatagcgt	atattaaagt	tgctgcagtt	aaaaagctcg	10320
tagttggatc	ttgggagcgg	gcgggcgggtc	cgcgcgagg	cgagtaccgc	cccgcccccg	10380
cccccttgcct	cctcgagccc	cctcgatgct	cttagctgag	tgtcccgcgg	ggcccgaagc	10440
gtttactttg	aaaaaattag	agtgttcaaa	gcaggcccga	gccgcctgga	taccgcagct	10500
aggaataatg	gaataggacc	gcggttctat	ttgttggtt	ttcggaactg	aggccatgat	10560
taagagggac	ggccgggggc	attcgtattg	cgcgcctaga	ggtgaaattc	ttggaccggc	10620

gcaagacgga	ccagagcgaa	agcatttgcc	aagaatgttt	tcattaatca	agaacgaaag	10680
tcggagggttc	gaagacgatc	agataccgtc	gtagttccga	ccataaaacga	tgccgactgg	10740
cgatgcccgcg	gcgttattcc	catgaccgcg	cgggcagctt	ccgggaaacc	aaagtctttg	10800
gggtccggggg	ggagtatggt	tgcaaagctg	aaacttaaaag	gaattgacgg	aaggggacca	10860
ccaggagtgg	gcctgcggct	taatttgact	caacacggga	aacctcacc	ggcccgga	10920
cggacaggtat	tgacagattg	atagctcttt	ctcgattccg	tggttggtgg	tgcatggccg	10980
ttcttagttg	gtggagcgat	ttgtctggtt	aattccgata	acgaacgaga	ctctggcatg	11040
ctaactagtt	acgcgacccc	cgagcggctg	gcgtccccc	acttcttaga	gggacaagtg	11100
gcgttcagcc	acccgagatt	gagcaataac	aggtctgtga	tgcccttaga	tgccggggg	11160
tgacgcgcg	ctacactgac	tggtcagcg	tggtccctacc	ctgcgcggc	aggcgcgggt	11220
aaccggtga	accccatctg	tgatggggat	cggggattgc	aattattccc	catgaacgag	11280
gaattcccag	taagtgcggg	tcataagctt	gcgttgatta	agtccctgcc	ctttgtacac	11340
accgcccgtc	gctactaccg	attggatggt	ttagtggagg	cctcggatcg	gccccgcgg	11400
ggtcggccca	cggccctggc	ggagcgtga	gaagacggtc	gaacttgact	atctagagga	11460
agtaaaagtc	gtacaaaggt	ttcgtagggt	gaacctgcgg	aaaggatcatt	aaacgggaga	11520
ctgtggagga	gcggcgggct	ggcccgcctc	ccccgtcttg	tgtgtgtcct	cgccgggagg	11580
cgctgcgtc	ccgggtcccc	tcgcccgcgt	gtggagcag	gtgtctggag	tgaggtgaga	11640
gaaggggtgg	gtggggctcg	tcggggtccg	ctctgggacc	ccctcgattt	cccccccc	11700
tcctctctcc	ctcgtccggc	tcgtacctcg	ccaccctacc	gcggcgggcg	ctgctcgcgg	11760
gcgtcttgcc	tccttcccg	ccggtctctc	cgtgtctacg	aggggcggta	cgctcgttacg	11820
ggttttccag	ccgtcccggg	ggcgttcgggt	cgctggggcg	cgcgctttgc	tcctccggca	11880
cccatccccg	ccgcggctct	ggcttttcta	cgttggtggg	ggcgggtgtc	gcgtgtgggg	11940
ggatgtgagt	gtcgcgtgtg	ggctcggccc	tcctcgatgc	acgcttttct	ggcctcgcgt	12000
gtcctccccg	ctcctgtccc	gggtacctag	ctgtcgcgtt	ccggcgcgga	ggtttaagga	12060
ccccgggggg	gtcgcctctg	cgccccccag	gtcggggggc	ggtggggccc	gtagggaaagt	12120
cgggtcgttcg	ggcgggtctc	cctcagactc	catgaccctc	ctccccccgc	tgccgcggtt	12180
cccgaggcgg	cgggtcgtgtg	gggggggtga	tgcttgagag	ccctcgggg	gcccgtgggg	12240
ccccgacccgc	gcccgcggct	tgcccgatth	ccgcgggtcg	gtcctgtcgg	tgccgggtcgt	12300
gggttcccgt	gtcgttcccc	tgtttttccg	ctcccgacc	tttttttttc	ctcccccca	12360
cacgtgtctc	gtttcgttcc	tgtggccgg	cctcaggcta	ccctcgggtc	catctgttct	12420
cctctctctc	cggggagagg	agggcggtgg	tcgttggggg	actgtgccgt	cgctcagacc	12480
cgtgagttcg	ctcacaccgg	aaataccgat	acgactctta	gcgggtggatc	actcggctcg	12540
tgctgcagat	aagaacgcag	ctagctgcga	gaatttaagt	gaattgcagg	acacattgat	12600
catcgacact	tcgaacgcac	ttgcggcccc	gggttccctc	cggggctacg	cctgtctgag	12660
cgtcggttga	cgatcaatcg	cgtcacccgc	tgccgtgggt	gctgcgcggc	tgggagtttg	12720
ctcgacggcg	caacccccca	acccgggtcg	ggccctccgt	ctcccgaaat	tcagacgtgt	12780
gggcgggtgt	cgggtgtggcg	cgccgcggcg	cgtcgcggag	cctgggtctc	cccgcgcac	12840
cgcgtcgcg	gcttcttccc	gctccggcgt	tcccgcctc	gcccgtgcac	cccgggtcgt	12900
gcctcgcgtc	ggcgccctcc	ggaccgctcg	ctcaccagtc	tttctcggtc	cogtcccccg	12960
tggaacccca	cccgcccccc	gtggcgcccg	gggggtggcg	cgtccgcac	tgctctgggtc	13020
gaagttggcg	gttgagggtg	tgctgcgcgc	gaggtggtgg	tcgggtcccct	gcggcgccgg	13080
gggtgtcggg	gtggcggtcg	acgaggccgc	gtcggtcgcc	tgccgtgggt	gtctgtgtgt	13140
gtttgggtct	tgccgtgggg	gaggcggggt	cgaccgctcg	cgggggttggc	cggtgcgcc	13200
ggcgccgcgc	accctccggc	ttgtgtggag	ggagagcgag	ggcgagaacg	gagagaggtg	13260
gtatccccgg	tgccgttggc	agggagggtt	tgccgtcccg	cgtccgtccg	tccttccctc	13320
cctcgggtggg	cgccttcggc	ccgcacgcgg	ccgttagggg	cggtcggggc	cgtggcccc	13380
cgtggctctt	cttcgtctcc	gcttctcctt	caccggggcg	gtaccgcctc	cggcgccggc	13440
ccgcgggacg	ccgcggcgct	cgtgcgcga	tgccagtcac	ccccgggtgt	tgccagttctg	13500
gggaggggaga	gggcctcgct	gacccgttgc	gtcccggtt	ccctggggggg	gacccggcgt	13560
ctgtgggctg	tgccgtccccg	gggttgccgt	tgagtaagat	cctccacccc	cgccgcctc	13620
ccctcccgcc	ggcctctcgg	ggacccctcg	agacggttcg	ccggctcgtc	ctcccggtgc	13680
gcccgggtgcc	gtctctttcc	cgcgcgcctc	ctcgtctctc	tcttcccgcg	gctgggcgg	13740
tgtccccctt	ttctgaccgc	gacctcagat	cagacgtggc	gacccgctga	atttaagcat	13800
attagtcagc	ggaggaaaag	aaactaacca	ggattccctc	agtaacggcg	agtgaacagg	13860
gaagagccca	gcgcggaatc	ccgcgcgcgc	gtcgcggcgt	gggaaatgtg	gcgtacggaa	13920
gacccaactcc	ccggcgccgc	tcgtgggggg	cccaagtcct	tctgatcgag	gcccagccc	13980
tggacgggtg	gaggccggta	gcggccccgg	cgcgcggggc	tcgggtcttc	ccggagtcgg	14040
gttgcttggg	aatgcagccc	aaagcgggtg	gtaaaactca	tctaaggcta	aatagccga	14100
cgagaccgat	agtcaacaag	taccgtaagg	gaaagtgtga	aagaactttg	aagagagagt	14160
tcaagagggc	gtgaaaccgt	taagaggtta	acgggtgggg	tccgcgcagt	ccgcccggag	14220
gattcaacc	ggcgccgcgc	gtccggcggt	cccggttggg	cccgccggat	ctttcccgct	14280
ccccgttcc	cccgacccct	ccacccgcgc	gtcgttcccc	tcttccctcc	cgcgtccggc	14340
gcctccggcg	gcgggcgcgg	gggggtggtg	ggtggtggcg	cgcggggcgg	gcccgggggtg	14400
gggtccggcg	gggacccccc	ccggccggcg	ccggccggcc	gcccggcgca	cttccaccgt	14460
ggcggtgcgc	cgcgaccggc	tcggggacgg	ccgggaaggc	ccgggtggga	aggtggctcg	14520
ggggggggcg	cgcgtctcag	ggcgcccgga	accacctcac	cccagagtgt	acagccctcc	14580
ggccgcgctt	tcgcggaatc	ccggggccga	ggaagccaga	taccgctcgc	cgcgtctctc	14640

ctctccccc	gtccgcctcc	cgggcgggcg	tgggggtggg	ggccggggcg	cccccccac	14700
ggcgcgaccg	ctctcccacc	ccctcccgtc	gcctctctcg	gggcccgggtg	ggggcgggg	14760
cggactgtcc	ccagtgcgcc	ccgggcgtcg	tcgcgcgcgtc	gggtcccggg	gggaccgtcg	14820
gtcacgcgtc	tcccgcagaa	gccgagcgca	cggggtcggc	ggcgatgtcg	gctaccacc	14880
cgaccgtct	tgaaacacgg	accaaggagt	ctaaccgctg	cgcgagtcag	gggctcgctcc	14940
gaaagccgcc	gtggcgcaat	gaaggtgaag	ggccccgcgc	gggggcccga	gggtgggtacc	15000
cgaggcctct	ccagtccgcc	gagggcgcac	caccggcccg	tctcgcccg	cgcgcgggg	15060
aggtggagca	cgagcgtacg	cgttaggacc	cgaagatgg	tgaactatgc	ttgggcagg	15120
cgaagccaga	ggaaactctg	gtggaggtcc	gtagcggtcc	tgacgtgcaa	atcggtcgtc	15180
cgacctgggt	atagggcgga	aagactaatc	gaaccatcta	gtagctgggt	ccctccgaag	15240
tttccctcag	gatagctggc	gctctcgctc	ccgacgtacg	cagttttatc	cggtaaagcg	15300
aatgattaga	ggctctgggg	ccgaaacgat	ctcaacctat	tctcaaactt	taaatgggtg	15360
agaagcccg	ctcgctggcg	tggagccggg	cgtggaatgc	gagtgcctag	tgggcccactt	15420
ttggtaagca	gaactggcgc	tgccggatga	accgaacgcc	gggttaaggc	gcccagtcg	15480
gacgctcate	agaccccaga	aaaggtgttg	gttgatatag	acagcaggac	gggtggcatg	15540
gaagtcggaa	tccgctaagg	agtgtgtaac	aactcacctg	ccgaatcaac	tagccctgaa	15600
aatggatggc	gctggagcgt	cgggcccata	cccgcccgctc	gccgcagtcg	gaacggaaag	15660
ggacggggagc	ggccgcggtg	gcgcgtctct	cggggtcggg	gggtgcgtggc	ggggcccgtg	15720
cccccgctc	ccctccgcgc	gccgggttcg	cccccgcggc	gtcgggcccc	gcccagacct	15780
cgccgcgacg	agtaggagg	ccgctgcggt	gagccttgaa	gcctaggggc	cgggcccggg	15840
tggagccgcc	gcaggtgcag	atcttggtgg	tagtagcaaa	tattcaaacg	agaaacttga	15900
aggccgaagt	ggagaagggt	tccatgtgaa	cagcagttga	acatgggtca	gtcggtcctg	15960
agagatgggc	gagtgccgtt	ccgaagggac	gggcgatggc	ctccgttgcc	ctcgcccgat	16020
tgaaaggag	tcgggttcag	atccccgaat	ccgagtggc	ggagatgggc	gcccgcaggc	16080
cagtgcggta	acgcgaccga	tcccgagaa	gccggcgggg	ggcctcgggg	agagttctct	16140
tttctttgtg	aaggccagg	cgccctggaa	tgggttcgcc	ccgagagagg	ggcccgtgcc	16200
ttgaaagcg	tcgcggttcc	ggcgcgctcc	ggtagctctc	cgctggccct	tgaaaatccg	16260
ggggagagg	tgtaaatctc	gcgcggggcc	gtaccatata	ccgcagcagg	tctccaagg	16320
gaacagcctc	tggcatgttg	gaacaatgta	ggtaagggaa	gtcggaagc	cggtaccgta	16380
acttcgggat	aaggattggc	tctaagggtc	gggtcggtcg	ggctggggcg	cgaagccggg	16440
ctgggcgcgc	gccgcggctg	gacgaggcgc	gccgcgcctc	tcccacgtcc	ggggagaccc	16500
cccgctcttt	ccgcgcgggc	ccgcctctcc	ctctcccccg	cggggccccg	tcgtcccccg	16560
cgctcgctcc	ccccctcttc	ccccctcttc	ctctcccgct	gggggcccgg	cgggggctcg	16620
cgcgcgccgc	gggctccggg	gcggcggggtc	caaccccgcg	gggggtccgg	agcgggagga	16680
accagcggtc	cccggtgggg	cgggggggccc	ggacactcag	ggggccggcg	gcccgcggcg	16740
ctctggacgc	gcgcggggcc	cttcccggtg	ctgcgcctcg	ctgcggcggg	cgtcgccggc	16800
gctcccgggg	agcccggggg	gtgcccggcg	gggtcccctc	cccgcggggc	ctcgctccac	16860
ccccccatcg	ccctctccga	gggtgcgtggc	ggggcgcggc	gggcgtgtcc	cgcgcgtgtg	16920
gggggaacct	ccgcgtcggt	gttccccgcg	cggttcgcgc	ccccgggccc	cggttttccg	16980
cgcgcgccgc	ccgcctcggc	cggcgcctag	cagccgaact	agaactgggt	cggaccagg	17040
gaatccgact	gtttaattaa	aacaaagcat	cgcgaaggcc	cgcggcgggg	gttgacgcga	17100
tgtgatttct	tccactgtct	ctgaatgtca	aagtgaagaa	attcaatgaa	gcgcgggtaa	17160
acggcgggag	ttaactatgac	tctcttaagg	tagccaaatg	ctcgtctatc	taattagtga	17220
cgcgcatgaa	tggatgaacg	agattccac	tgtccctacc	tactatccag	cgaaccaca	17280
gccaagggaa	cgggcttggc	ggaatcagcg	gggaaagaag	accctgttga	gcttgactct	17340
agtctggcac	gggtgaagaga	catgagaggt	gtagaataag	tgggaggccc	ccggcgcccg	17400
gccccgtcct	cgcgctcggg	tcggggcacg	ccggcctcgc	gggcgcggcg	tgaataacca	17460
ctactctcat	cggttttttca	ctgaccgggt	gaggcggggg	ggcgagcccc	gaggggctct	17520
cgcttctggc	gccaagcgtc	cgctccgcgc	gtgcggggcg	gcgcgacccg	ctccggggac	17580
agtgcagggt	ggggagtttg	actggggcg	tacacctgtc	aaacggtaac	gcaggtgtcc	17640
taaggcgagc	tcaggggagga	cagaaacctc	ccgtggagca	gaagggcaaa	agctcgcttg	17700
atcttgattt	tcagtacgaa	tacagaccgt	gaaagccggg	cctcacgata	cttctgacct	17760
tttgggtttt	aagcaggagg	tgtcagaaaa	gttaccacag	ggataactgg	cttgtggcg	17820
ccaagcgttc	atagcgacgt	cgcttttttga	tccttcgatg	tcggctcttc	ctatcattgt	17880
gaagcagaat	tcaccaagcg	ttggattgtt	caccactaa	tagggaacgt	gagctgggtt	17940
tagaccgtcg	tgagaagag	tagttttacc	ctactgatga	tgtgttgttg	ccatggtaat	18000
cctgctcagt	acgagaggaa	ccgcagggtt	agacatttgg	tgtatgtgct	tggctgagga	18060
gccaatgggg	tcacacagcg	atctgtggga	acgcctctaa	acgcctctaa	gtcagaatcc	18120
gccaagcgg	aacgatacgg	cagcgccgaa	ggagcctcgg	ttggccccgg	atagccgggt	18180
ccccgtccgt	cccgctcggc	ggggcccccg	cgctgcggcg	cggcggcgcg	gggtctcccc	18240
ccgcggggcg	tcgggacggg	ggctccggtg	ggctccggtg	tcgtcttggg	aaacgggggtg	18300
cgcccggaaa	gggggcccgc	ctctcgcccg	tcacgttgaa	cgcacgttcg	tgtggaacct	18360
ggcgctaaac	cattcgtaga	cgacctgctt	ctgggtcggg	gtttcgtagc	tagcagagca	18420
gctccctcgc	tcgatctat	tgaaagtccg	ccctcgacac	aagggtttgt	ctctcgggc	18480
tttcccgctg	cacgcccgt	cgctcgacg	cgacctgtc	gccgcgggg	cgtcacgggg	18540
gcggctcgct	cgcccccg	gcgggtgccc	gaacgacgt	gtgggtgggt	ggggggggat	18600
cgtcttctcc	tcctctctcc	gaggacgggt	cgtttctctt	tccctctccg	tcgctctcct	18660

-9-

tgggtgtggg	agcctcgtgc	cgtcgcgacc	gccccctgcc	gtcgcctgcc	gcccgcagccc	18720
cttgccctcc	ggccttggcc	aagccggagg	gcggaggagg	gggatcggcg	gcggcgggcga	18780
ccgcggcgcg	gtgacgcacg	gtgggatccc	catcctcggc	gcgtccgtcg	gggacggccg	18840
gttggagggg	cgggaggggt	ttttcccgtg	aacgcgcgt	tcggcgccag	gcctctggcg	18900
gccggggggg	cgctctctcc	gcccgagcat	ccccactccc	gccccctctc	ttcgcgcgcc	18960
gcggcgggcga	cgtgcgtacg	aggggaggat	gtcgcgggtg	ggaggcgagg	aggggtccggc	19020
gcggcgccctc	ttccattttt	ttccccccaa	cttcggagggt	cgaccagtac	ttccggcgagc	19080
actttgtttt	ttttttttcc	cccgatgctg	gaggtcgacc	agatgtccga	aagtgtcccc	19140
cccccccccc	ccccccggcg	cggagcgggc	ggggccactct	ggactctttt	tttttttttt	19200
tttttttttt	ttaaattcct	ggaaccttta	ggtcgaccag	ttgtccgtct	tttactcctt	19260
catataggtc	gaccagtact	ccgggtggta	ctttgtcttt	ttctgaaaat	cccagagggtc	19320
gaccagatat	ccgaaagtcc	ttctttttcc	tttactcttc	cccacagcga	ttctcttttt	19380
tttttttttt	tttgggtgtg	ctcttttttg	cttatataca	tgtaaatagt	gtgtacgttt	19440
ataacttat	aggaggaggt	cgaccagtac	tcggggcgac	actttgtttt	tttttttttt	19500
ttccacgatg	atggagggtg	accagatgtc	cgaaagtgtc	ccgtcccccc	ctcccccccc	19560
ccgcgacgcg	gcgggctcac	ttctggactct	tttttttttt	tttttttttt	tttaaatttc	19620
tggaacctta	aggctcgacca	gttgtccgtc	tttctactcat	tcatataggt	cgaccgggtgg	19680
tactttgtct	ttttctgaaa	atcgagagg	tcgaccagat	gtcgaaaagt	ctgggtggctg	19740
ataaattatc	tgatctagat	ttgtttttct	gtttttcagt	tttgtgttgt	tttgtgttgt	19800
tttgtgttgt	tttgttttgt	tttgttttgt	tttgttttgt	tttgttttgt	tttgttttgt	19860
tttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	19920
gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	19980
tttgtgttgt	tttgttttgt	tttgttttgt	tttgttttgt	tttgttttgt	tttgttttgt	20040
tacacaaaca	tgacattttt	ttaaaataaa	tttttaaaat	aaatgcgaaa	atcgaccaat	20100
tatccctttc	cttctctctc	ttttttaaaa	attttctttg	tgtgtgtgtg	tgtgtgtgtg	20160
tgtgtgtgtg	tgctgtgtgt	tgctgtgtgt	cgctgcagct	gcgcgcgctc	gttttataaa	20220
tacttataat	aataggtcgc	cggtgtgtgt	tagcttccc	gactccagag	gcagaggcg	20280
gcagacttct	gagttcgagg	ccagcctggg	ctacagagga	accctgtctc	gaaaaatgaa	20340
aataaataca	tacatacata	catacataca	tacatacata	catacataca	tacatatgag	20400
gttgaccagt	gtcaatcct	ttagaatttt	gtttttaatt	aatgtgatag	agagatagat	20460
aatagataga	tggatagagt	gatacaata	taggtttttt	tttcagtaaa	tatgaggttg	20520
attaaccact	tttccctttt	taggtttttt	tttttttccc	ctgtccatgt	gggtgtgtgg	20580
atttgaactc	tgacacctgg	caggtcaact	ggaaaacgtg	ttttctatat	atataaatag	20640
tgggtctgtct	gctgttttgt	tggtttgctg	cttgcttgct	tgcttgcttg	cttgcttgct	20700
tgtttttttt	tttcttctga	gacagtat	ctctgtgtaa	cctgggtgcc	tgaaactcac	20760
ctgttagacc	agcctggcct	caatcgaa	cagaaatcct	cctgcctctt	gtctacctcc	20820
caattttgga	gtaaagggtg	gctacaccac	tgcttggtat	tattatcatt	atcattatta	20880
atatttattt	tagacagaac	gaaatcaact	agttgggtct	gtttcggtta	ttcatttgaa	20940
atttagttga	ccaattagtt	ggctgggttt	ggagggtttt	tttgtttccg	atttgggtgt	21000
tttgtgggct	ggggatcagg	tatctcaacg	gaatgcattg	agggttaagg	gagatggctc	21060
gattttttga	aagattactt	ttcttagtct	gaggaaaaaa	taaaataata	ttgggctacg	21120
tttcatttgt	ctatttctat	ttctctttct	ttctttcttt	ctttcagata	aggaggctcg	21180
ccagttcctc	ctgccttctg	gaagatgtag	gcattgcatt	gggaaaagca	ttgtttgaga	21240
gatgtgctat	tgaaccagag	agtttggatg	tcaagccgta	taattgtttat	tacaatatag	21300
aaaagtctta	acaaagtgat	ctttaacttt	tttttttttt	tttctccttc	tacttctact	21360
tgttctcact	ctgccacca	cgcgctttgt	acattgaaat	tgagctttgt	tttgcttaac	21420
agacatatat	tttttctttt	ggttttgctt	gacatgggtt	ccctttctat	ccgtgcaggg	21480
ttcccagacg	gccttttgag	aataaaatgg	gaggccagaa	ccaaagtctt	ttgaataaag	21540
caccacaact	ctaaccgtgt	tggtgttttt	ccttcccaag	gcacagatct	ttccagcat	21600
ggaaaagcat	gtagcagttg	taggacacac	tagacgagag	caccagatct	cattgtgggt	21660
ggttgtgaac	caccaccat	gtggttgcct	gggatttgaa	ctcaggatct	tcagaagacg	21720
agtcagggtc	ctaaaccgat	gagccatctc	tccagccctc	ctacattcct	tcttaaggca	21780
tgaatgatcc	cagcatggga	agacagtctg	ccctctttgt	ggtatatcac	catatactca	21840
ataaaataat	gaaatgaatg	aagtctccac	gtatttattt	cttcgagcta	tctaaattct	21900
ctcacagcac	ctccccctcc	cccacactgc	ctttctctct	atgtttgggt	ggggctgggg	21960
gaggggtggg	gtgggggag	ggatctgcac	gtcttcttgc	aggtctgtga	actatttgcg	22020
atggcctggg	tctctgaact	gttgagcctt	gtctatccag	aggtgactg	gctagttttc	22080
tacctgaagt	ccctgagtga	tgatttccct	gtgaattc			22118

<210> 19

<211> 175

<212> DNA

<213> Mus musculus

<400> 19

ctccccgcgc	gcccccggtg	tcgcgcgttc	cgtggcgcg	acaatgcggt	tgtgcgtcca	60
cgtgtgcgtg	tcggtgcagt	gccgttgtgg	agtgccctgc	tctcctctc	ctccccggca	120

-10-

gcgttcccac gggtggggac caccgggtgac ctgcacctct tcgggcctgg atccg 175

<210> 20
 <211> 755
 <212> DNA
 <213> Mus musculus

<400> 20
 ggtctggtgg gaattgttga cctcgctctc ggggtgcggcc tttggggaac ggcgggggtcg 60
 gtctgccccg gcgcgggacg tgtgtcgggg cccacttccc gctcgagggg ggcgggtggcg 120
 gcggcggttg tagtctcccc tgttgctctc tcccggggctc ttgggggggg tgccgtcgtt 180
 ttccggggccg gcgttgcttg gcttacgcag gcttggtttg ggactgcctc aggagtcgtg 240
 ggcggtgtga ttcccggcg ttttgccctcg cgtctgcctg ctttgccctc ggtttgcttg 300
 gttcgtgtct cgggagcggg ggtttttttt ttttcgggtt cccggggaga ggggtttttc 360
 cgggggacgt tcccgtcgcc ccctgccgcc ggtgggtttt cgtttcgggc tgtgttcggt 420
 tccccctccc cgtttcggcg tcggtttctcc ccggtcggtc ggccctctcc ccggtcggtc 480
 gcccggccgt gctgccggac ccccccttct ggggggggatg cccgggcacg cacgcgtccg 540
 ggcggccact gtggtccggg agctgctcgg caggcgggtg agccagttgg aggggcgtca 600
 tgcccccgcg ggtcccggtg gccgacggcg cgtgttcttt gggggggcct gtgcgtgcgg 660
 gaaggtctgc cacgttgtcg gtccttgcca gggaaagagg cttttttttt ttaggggggtc 720
 gtccttcggt gtcccgtcgg cgggtggatcc ggcct 755

<210> 21
 <211> 463
 <212> DNA
 <213> Mus musculus

<400> 21
 ggcgaggtg cgtctgcggg ttggggctcg tccggccccg tcgtcctccg ggaaggcgtt 60
 tagcgggtac cgtccgcgcg ccgaggtggg cgcacgtcgg tgagataacc ccgagcgtgt 120
 ttctggttgt tggcggcggg ggctccggtc gatgtcttcc cctccccctc tccccgagggc 180
 caggtcagcc tccgcctgtg ggcttcgtcg gccgtctccc cccccctcac gtccctcgcg 240
 agcagccccg tccgttcgac cttccttccg cctccccccc atctttccgc gtcctcgttg 300
 ccccggggtt ttcaaggcgc cccccacgct cctccgcctc tccgcccgtg gtttggacgc 360
 ctggttccgg tctccccgcc aaaccccggt tgggttggtc tccggccccg gcttgccttt 420
 cgggtctccc aacccccggc cgggaagggt cgggggttcc ggg 463

<210> 22
 <211> 378
 <212> DNA
 <213> Mus musculus

<400> 22
 ggattcttca ggattgaaac ccaaaccggt tcagtttctt ttccggctcc ggcggggggg 60
 ggcggccccg ggcgggtttg tgagttagat aacctcgggc cgatcgcacg ccccccgtag 120
 cggcgacgac ccattcgaac gtctgcccct tcaactttcg atggtagtcg atgtgcctac 180
 catggtgacc acgggtgacg ggggaatcagg gttcgattcc ggagagggag cctgagaaac 240
 ggctaccaca tccaaggaag gcagcaggcg cgcaaattac ccactcccga cccggggagg 300
 tagtgacgaa aaataacaat acaggactct ttcgaggccc tgtaattgga atgagtcac 360
 tttaaatcct ttaagcag 378

<210> 23
 <211> 378
 <212> DNA
 <213> Mus musculus

<400> 23
 gatccattgg agggcaagtc tggtgccagc agccgcggta attccagctc caatagcgta 60
 tattaaagtt gctgcagtta aaaagctcgt agttggatct tgggagcggg cgggcggttc 120
 gccgcgaggg gagtcaccgc ccgtccccgc ccttgccctc tcggcgcccc ctcgatgctc 180
 ttactgagtt tgtcccgagg ggcccggaag gtttactttg aaaaaattag agttgtttca 240
 aagcaggccc gagccgcctg gataccgcca gctaggaaat aatggaatag gaccgcggtt 300
 cctattttgt ttggttttcg gaactgagcc catgattaag ggaaacggcc gggggcattc 360
 ccttattgag ccccccta 378

<210> 24
 <211> 719

-11-

<212> DNA

<213> Mus musculus

<400> 24

ggatctttcc	cgctccccgt	tcttccccgc	ccctccaccc	gcgcgtctcc	ccccttcttt	60
tccccctctcc	ggaggggggg	gaggtggggg	cgctggggcg	gggtcggggg	tggggtcggc	120
gggggaccgc	ccccggcccg	caaaaggccg	ccgccgggcg	cacttcaacc	gtagcgggtg	180
gcgcgacccg	gctacgagac	ggctgggaag	gcccgaacgg	gaatgtggct	cggggggggc	240
ggcgcgctctc	agggcgcgcc	gaaccacctc	accccgagtg	ttacagccct	ccggcgcgcg	300
tttcgcgga	tcccggggcc	gaggggaagc	ccgatacccc	tgcgcgcgct	tttccccctc	360
ccccgtccgc	ctcccgggcg	ggcgtggggg	tggggggcgg	gccgccccctc	ccacgcctgt	420
ggttttctctc	tctcccggtc	tgcggccggtt	tggggggggg	agcccgggtg	ggggcggggg	480
ggactgtcct	cagtgcgccc	cgggcgtcgt	cgccgcgtcg	ggcccggggg	gttctctcgg	540
tcacgcgcgc	cccgaacga	ccgagcgcac	gggtcgggcg	gcgatgtcgg	ctacccaccc	600
gacccgtctt	gaaacacgga	ccaaggagtc	taacgcgtgc	gcgagtcagg	ggctcgcaag	660
aaagccgccc	tggcgcaatg	aaggtgaagg	gccccgtccg	ggggcccag	gtgggatcc	719

<210> 25

<211> 685

<212> DNA

<213> Mus musculus

<400> 25

cgaggcctct	ccagtcgcgc	gagggcgcac	caccggcccc	tctcgccccg	cgcgctgggg	60
aggtggagca	cgagcgtacg	cgttaggacc	cgaaagatgg	tgaactatgc	ctgggagagg	120
cgaagccaga	ggaaaactctg	gtggaggtcc	gtagcgggtc	tgacgtgcaa	atcggctcgc	180
cgacctgggt	atagggggcg	aagactaatc	gaaccatcta	gtagctgggt	ccctccgaag	240
tttccctcag	gatagctggc	gctctcgcaa	ccttcggaag	cagttttatc	cgggttaaagg	300
cggaatggat	taggaggtct	tggggccgga	aacgatctca	aactatttct	caaactttaa	360
atgggtaagg	aagccccgct	cgctggcggtg	gagccgggcg	tggaaatgcga	gtgcctagtg	420
ggccactttt	ggtaagcaga	actggcgctg	cgggatgaac	cgaacgcccg	gttaaggcgc	480
ccgatgccga	cgctcatcag	accccagaaa	aggtgttggt	tgatatagac	agcaggacgg	540
tggccatgga	agtcggaatc	cgctaaggag	tgtgtaacaa	ctcacctgcc	gaatcaacta	600
gccctgaaaa	tggatggcgc	tggagcgctg	ggcccatacc	cggccgtcgc	cggcagtcgg	660
aacgggacgg	gacgggagcg	gccgc				685

<210> 26

<211> 5162

<212> DNA

<213> Artificial Sequence

<220>

<223> Chimeric bacterial plasmid

<400> 26

gacggatcgg	gagatctccc	gatccccat	ggtcgactct	cagtacaatc	tgctctgatg	60
ccgcatagtt	aagccagtat	ctgctccctg	cttgtgtgtt	ggaggtcgct	gagtagtcgg	120
cgagcaaat	ttaagctaca	acaaggcaag	gcttgaccga	caattgcatg	aagaatctgc	180
ttagggttag	gcgttttgcg	ctgcttcgcg	atgtacgggc	cagatatacg	cgttgacatt	240
gattattgac	tagttattaa	tagtaataca	ttacggggtc	attagttcat	agcccatata	300
tggagttccg	cgttacataa	cttacggtaa	atggcccgcc	tggtgaccg	cccaacgacc	360
ccgcccatt	gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	420
attgacgtca	atgggtggac	tattttacgtt	aaactgcccc	cttggcagta	catcaagtgt	480
atcatatgcc	aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	540
atgcccagta	catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtc	600
tgcctattac	attgggtgatg	cggtttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	660
actcacgggg	atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	720
aaaatcaacg	ggactttcca	aaatgtcgta	acaactccgc	ccatttgacg	caaattggcg	780
gtaggcggtg	acgggtgggag	gtctatataa	gcagagctct	ctggctaact	agagaaccca	840
ctgcttactg	gcttatcgaa	attaatacga	ctcactatag	ggagacccaa	gcttgggtacc	900
gatctcggat	cgatatctgc	ggccgcgctg	acggaattca	gtggatccac	tagtaacggc	960
cgccagtggt	ctggaattaa	ttcgctgtct	gcgagggcca	gctgttgggg	tgagtactcc	1020
ctctcaaaag	cgggcatgac	ttctgcgcta	agattgtcag	tttccaaaaa	cgaggaggat	1080
ttgatattca	cctggcccg	ggtgatgcct	ttgaggttgg	ccgcgtccat	ctggtcagaa	1140
aagacaatct	ttttgttgte	aagcttgagg	tgtggcaggc	ttgagatctg	gccatacact	1200
tgagtgaaca	tgacatccac	tttgcccttc	tctccacagg	tgtccactcc	caggtccaac	1260
tgcaggtcga	gcatgcatct	agggcgggcca	attccgcccc	tctccctccc	ccccccctaa	1320

-12-

cggttactggc	cgaagccgct	tgggaataagg	cgggtgtgcy	tttgtctata	tgtgattttc	1380
caccatattg	ccgtcttttg	gcaatgtgag	ggcccggaag	cctggccctg	tcttcttgac	1440
gagcattcct	aggggtcttt	ccctctcgc	caaaggaatg	caaggtctgt	tgaatgtcgt	1500
gaaggaagca	gttctctctg	aagcttcttg	aagacaaaca	acgtctgtag	cgaccctttg	1560
caggcagcgg	aaccccccc	ctggcgacag	gtgcctctgc	ggccaaaagc	cacgtgtata	1620
agatacacct	gcaaaggcgg	cacaacccca	gtgccacgtt	gtgagttgga	tagttgtgga	1680
aagagtcaaa	tggctctcct	caagcgtatt	caacaagggg	ctgaaggatg	cccagaaggt	1740
accccatgtg	atgggatctg	atctggggcc	tgggtgcaca	tgctttacat	gtgttttagt	1800
gaggttaaaa	aaacgtctag	gccccccgaa	ccacggggac	gtggttttcc	tttgaaaaac	1860
acgatgataa	gcttgccaca	accggggatc	caccgggtcg	caccatgggt	agcaagggcg	1920
aggagctggt	caccgggggt	gtgcccaccc	tggtcgagct	ggacggcgac	gtaaacggcc	1980
acaagttcag	cgtgtccggc	gagggcgagg	gcgatgccac	ctacggcaag	ctgaccctga	2040
agttcatctg	caccaccggc	aagctgcccg	tgccctggcc	caccctcgtg	accaccctga	2100
cctacggcgt	gcagtgcttc	agcgcctacc	ccgaccacat	gaagcagcac	gacttcttca	2160
agtcgcctat	gcccgaagge	tacgtccagg	agcgacccat	cttcttcaag	gacgacggca	2220
actacaagac	ccgcgcggag	gtgaagttcg	agggcgacac	cctggtgaac	cgcacgcagc	2280
tgaagggcat	cgacttcaag	gaggacggga	acatcctggg	gcacaagctg	gagtagaact	2340
acaacagcca	aaacgtctat	atcatggcgg	acaaggggaa	gaacggcatc	aaggtgaact	2400
tcaagatccg	ccacaacatc	gaggacggga	gcgtgcagct	cgccgaccac	taccagcaga	2460
acacccccat	cggcgacggc	cccggtgctg	tgcccgacaa	ccactacctg	agcaccaggt	2520
ccgcccctag	caaagacccc	aacgagaagc	gcgatcacat	ggtcctgctg	gagttcgtga	2580
ccgcgcgcgg	gatcactctc	ggcatggacg	agctgtacaa	gtaaagcggc	cctagagctc	2640
gctgatcagc	ctcgactgtg	cctctagtgt	ccagccatct	gttggttgcc	cctccccctg	2700
gccttctctg	accctggaag	gtgcccactc	cactgtcctt	tcctaataaa	atgaggaaat	2760
tgcacgcgat	tgtctgagta	gggtgctatc	tattctgggg	ggtgggggtg	ggcaggacag	2820
caagggggag	gattgggaag	acaatagcag	gcctgtctgg	gatgcgggtg	gctctatggc	2880
ttctgaggcg	gaaagaacca	gctggggctc	gagtgcatcc	tagttgtggg	ttgtcccaac	2940
tcatacaatg	atcttatcat	gtctgtatcc	cgctgcacct	tagctagagc	ttggcgtaat	3000
catggtcata	gctgtttcct	gtgtgaaatt	gttatccgct	cacaattcca	cacaacatac	3060
gagccggaag	cataaagtgt	aaagcctggg	gtgcctaatt	agttagctaa	ctcacattaa	3120
ttgcgttgog	ctcactgccc	gctttccagt	cgggaaacct	gtcgtgccag	ctgcattaat	3180
gaatcggcca	acgcgcgggg	agaggcgggt	tgccgtattg	gcgctcttcc	gcttccctgc	3240
tcactgactc	gctgcccctg	gtcgttcggc	tgccggcgag	ggtatcagct	cactcaaagg	3300
cggtaataacg	gttatccaca	gaatcagggg	ataacgcagg	aaagaacctg	tgagcaaaag	3360
gccagcaaaa	ggccaggaac	cgtaaaaagg	ccgcgtttgt	ggcgtttttc	cataggctcc	3420
gcccccttga	cgagcatcac	aaaaatcgac	gctcaagtca	gaggtggcga	aacccgcagc	3480
gactataaag	ataccaggcg	tttccccctg	gaagctccct	cgtgcgctct	cctgttccga	3540
cctgcccgtc	taccggatcc	ctgtccgcct	ttctcccttc	gggaagcgtg	gcgctttctc	3600
aatgctcacc	ctgtaggtat	ctcagttcgg	tgtaggtcgt	tcgctccaag	ctgggctgtg	3660
tgcacgaacc	ccccgttcag	cccgacccgt	gcgccttctc	cggttaactat	cgcttgaggt	3720
ccaacccggg	aagacacgac	ttatcgccac	tggcagcagc	cactggtaac	aggattagca	3780
gagcgaggta	tgtaggcggg	gctacagagt	tcttgaagtg	gtggcctaac	tacggctaca	3840
ctagaaggac	agtatttggg	atctgcgcct	tgttgaagcc	agttaccttc	ggaaaaagag	3900
ttggtagctc	ttgatccggc	aaacaaacca	ccgctggtag	cggtgggttt	tttgtttgca	3960
agcagcagat	tacgcgcaga	aaaaaaggat	ctcaagaaga	tcctttgatc	ttttctacgg	4020
ggctctgacg	tcagtgggac	gaaaactcac	gttaagggat	tttgggtcat	agattatcaa	4080
aaaggatctt	cacctagatc	cttttaaatt	aaaaatgaag	ttttaaatca	atctaaagta	4140
tatatgagta	aacttgggtc	gacagttacc	aatgcttaat	cagttaggca	cctatctcag	4200
cgatctgtct	atttcgttca	tcocatagtt	cctgactccc	cgctgtgtag	ataactacga	4260
tacgggaggg	cttaccatct	ggccccagtg	ctgcaatgat	accgcgagac	ccacgctcac	4320
cggtctccaga	tttatcagca	ataaacccag	cagccgggag	ggccgagcgc	agaagtgggt	4380
ctgcaacttt	atccgcctcc	atccagctca	tttaattgtg	ccgggaagct	agagtaagta	4440
gttcgccaggt	taatagtttg	cgcaacgttg	ttgccattgc	tacaggcatc	gtgggtgtcac	4500
gctcgtcggt	tggtatgggt	tcattcagct	ccggttccca	acgatcaagg	cgagttacat	4560
gateccccat	gttgtgcaaa	aaagcgggta	gctccttcgg	tcctccgata	gttgtcagaa	4620
gtaagttggc	cgcagtggtt	tcactcatgg	ttatggcagc	actgcataat	tctcttactg	4680
tcattgccatc	cgtaagatgc	ttttctgtga	ctgggtgagta	ctcaaccaag	tcattctgag	4740
aatagtgatg	cgccgcagcg	agttgctctt	gcccgcgctc	aatacgggat	aataccgcgc	4800
cacatagcag	aactttaaaa	gtgctcatca	ttggaaaacg	ttcttcgggg	cgaaaactct	4860
caaggatctt	accgctgttg	agatccagtt	cgatgtaaac	cactcgtgca	cccaactgat	4920
cttcagcatc	ttttactttc	accagcgttt	ctgggtgagc	aaaaacagga	aggcaaatg	4980
ccgcaaaaaa	gggaataaagg	gcgacacgga	aatgttgaat	actcatactc	ttcctttttc	5040
aatattattg	aagcattttat	caggggttatt	gtctcatgag	cggatacata	tttgaatgta	5100
tttagaaaaa	taaacaaata	gggggtccgc	gcacattttc	ccgaaaagtg	ccacctgacg	5160
tc						5162

-13-

<211> 5627

<212> DNA

<213> Artificial Sequence

<220>

<223> pMG plasmid from InvivoGen; IRES sequence modified
EMCV nucleotides 2736-3308

<400> 27

caccggcgaa	ggaggcctag	atctatcgat	tgtacagcta	gctcgacatg	ataagataca	60
ttgatgagtt	tggacaaaacc	acaactagaa	tgcagtga	aaaatgcttt	atttgtgaaa	120
tttgtgatgc	tattgcttta	tttgtgaaat	tttgtgatgc	attgctttat	ttgtaaccat	180
tataagctgc	aataaacaag	tttaacaaca	caattgcatt	catttttatgt	ttcagggtca	240
gggggaggtg	tgggaggttt	tttaaagcaa	gtaaaacctc	tacaaatgtg	gtagatccat	300
ttaaatgtta	attaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	360
ggcgcggttg	ctggcggttt	tccataggct	ccgccccctc	gacgagcatc	acaaaaatcg	420
acgctcaagt	cagagggtgc	gaaacccgac	aggactataa	agataaccag	cgtttcccc	480
tggaaagctcc	ctcgtgcgct	ctcctgttcc	gaccctgcgc	cttaccggat	acctgtccgc	540
ctttctccct	tccgggaagcg	tggcgctttc	tcatagctca	cgctgtagg	atctcagttc	600
gggtgaggte	tgtcgctcca	agctgggctg	tgtgcacgaa	cccccgctc	agcccgaccg	660
ctgcgcctta	tccggtaact	atcgtcttga	gtccaacccg	gtaagacacg	acctatcgcc	720
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	780
gttcttgaag	tgggtgccta	actacggcta	cactagaaga	acagtatttg	gtatctgcgc	840
tctgctgaag	ccagttacct	tccgaaaaag	agttggtagc	tcttgatccg	gcaaaacaa	900
caccgctggt	agcgggtggt	ttttgttttg	caagcagcag	attacgcgca	gaaaaaaagg	960
atctcaagaa	gattccttga	tctttctctc	ggggtctgac	gctcagtggg	acgaaaaactc	1020
acgttaaggg	atcttgggtc	tggctagtta	attaagctgc	aataaacaat	cattatcttc	1080
attggatctg	tgtgttggtt	ttttgtgttg	gcttggggga	gggggaggcc	agaatgactc	1140
caagagctac	aggaaggcag	gtcagagacc	ccactggaca	aacagtggct	ggactctgca	1200
ccataacaca	caatcaacag	gggagtgcgc	tggatcgagc	tagagtcctg	tacataactt	1260
acggtaaatg	gcccgcctgg	ctgaccgcgc	aacgaccccc	gcccattgac	gtcaataatg	1320
acgtatgttc	ccatagtaac	gccaataggg	actttccatt	gacgtcaatg	ggtggagtag	1380
ttacggtaaa	ctgcccaact	ggcagtagat	caagtgtatc	atatgccaag	tacgccccct	1440
attgacgtca	atgacggtaa	atggccccgc	tggcattatg	cccagtacat	gaccttatgg	1500
gactttccta	cttggcagta	catctacgta	ttagtctatc	ctattaccat	ggtgatgcgg	1560
ttttggcagt	acatcaatgg	gcgtggatag	cggtttgact	cacggggatt	tccaagtctc	1620
caccccatgg	acgtcaatgg	gagtttgttt	tggcaccaaa	atcaacggga	ctttccaaaa	1680
tgtcgttaaca	actccgcccc	attgacgcaa	atggggcgga	ggcgtgtacg	gtgggaggtc	1740
tataaagca	gacgtcgttt	agtgaacggt	cagatcgcc	ggagacgcca	tccacgctgt	1800
tttgacctcc	atagaagaca	ccgggaccga	tccagcctcc	gcggccggga	acggtgcatt	1860
ggaacgcgga	ttccccgtgc	caagagtgc	gtaagtaccg	cctatagagt	ctataggccc	1920
accccttgg	cttcttatgc	atgctatact	gtttttggct	tggggtctat	acacccccgc	1980
ttcctcatgt	tataggtgat	ggtatagctt	agcctatagg	tgtgggttat	tgaccattat	2040
tgaccactcc	cctattgggt	acgatacttt	ccattactaa	tccataacat	ggctctttgc	2100
cacaactctc	tttattggct	atatgccaat	acactgtcct	tcagagactg	acacgggact	2160
tgtattttta	caggatgggg	tctcatttat	tatttacaaa	ttcacatata	caacaccacc	2220
gtccccagtg	cccgcagttt	ttattaaaca	taacgtggga	tctccacgcg	aatctcgggt	2280
acgtgttccg	gacatgggct	cttctccggg	agcggcgagg	cttctacatc	cgagccctgc	2340
tcccatgcct	ccagcgactc	atggctcgct	ggcagctcct	tgctcctaac	agtggaggcc	2400
agacttaggc	acagcacgat	gcccaccacc	accagtgtgc	cgcacaaggc	cgtggcggtg	2460
gggtatgtgt	ctgaaaatga	gctcggggag	cgggcttgca	ccgctgacgc	atttgggaaga	2520
cttaaggcag	cggcagaaga	agatgcaggc	agctgagttg	tttgtgtctg	ataagagtca	2580
gaggtaaact	ccgttgcggt	gctgttaacg	gtggagggca	gtgtagtctg	agcagtagtc	2640
gttgctgccc	cgcgcgccac	cagacataat	agctgacaga	ctaacagact	gttctcttcc	2700
atgggtcttt	tctgcagtca	cccgggggat	ccttcgaacg	tagctctaga	ttgagtcgac	2760
gttactggcc	gaagccgctt	ggaataaggc	cgggtgtcgt	ttgtctatat	gttattttcc	2820
accatattgc	cgtcttttgg	caatgtgagg	gcccggaac	ctggccctgt	cttcttgacg	2880
agcatttcta	gggggtcttt	ccctctcgcc	aaagggaatg	aagggtctgt	gaatgtcggt	2940
aagggaagcag	ttcctctgga	agcttcttga	agacaaacaa	cgtctgtagc	gaccttttgc	3000
aggcagcgga	acccccacc	tggcgacagg	tgccctctgc	gccaaaagcc	acgtgtataa	3060
gatacacctg	caaaggcggc	acaacccag	tgccacgttg	tgagttggat	agttgtggaa	3120
agagtcaaat	ggctctcctc	aagcgtatcc	aacaaggggc	tgaaggatgc	ccagaaagga	3180
ccccattgta	tgtgagctga	tctggggcct	gggtgacatg	tggttttctc	ttgaaaaaca	3240
aggttaaaaa	aacgtctagg	ccccccgaac	cacggggacg	ggcgccctagt	gttgaccatt	3300
cgataatacc	atgggtaagt	gatattctact	aattgtgacc	actataggag	ggccaccatg	3360
aatctatcgc	atagtatatc	ggcatagtag	aatacagctc	ccggtaggag	ggccatcatg	3420
tcgactacta	acctttctct	ctttctctaca	gctgagatca			3480

-14-

```

aaaaagcctg aactcacgcg gacgtctgtc gcgaagtttc tgatcgaaaa gttcgacagc 3540
gtctccgacc tgatgcagct ctccgagggc gaagaatctc gtgctttcag cttcgatgta 3600
ggagggcgctg gatatgtcct gcgggtaaat agctgcgcgc atggtttcta caaagatcgt 3660
tatgtttatc ggcacttttg atcggccgcg ctcccgattc cggaaagtgt tgacattggg 3720
gaattcagcg agagcctgac ctattgcatc tcccgcctgt cacagggtgt cacgttgcaa 3780
gacctgcctg aaaccgaact gcccgctgtt ctgcaaccgc tcgaggagct catggatgcg 3840
atcgctgcgg ccgatcttag ccagacgagc ggggttcggc cattcggacc gcaaggaatc 3900
ggccaataca ctacatggcg tgatttcata tgcgcgattg ctgatcccca tgtgtatcac 3960
tggcaaaactg tgatggacga caccgtcagt gcgtccgtcg cgcaggctct cgatgagctg 4020
atgctttggg ccgaggactg ccccgaaagt cggcacctcg tgcacgcgga tttcggctcc 4080
aacaatgtcc tgacggacaa tggccgcata acagcggcca ttgactggag cgaggcgatg 4140
ttcgggggatt cccaatacga ggtcgccaac atcttcttct ggaggccgtg gttggcttgt 4200
atggagcagc agacgcgcta cttcgagcgg aggcattccg agcttgacag atccgcgcgg 4260
ctccggcgct atatgctccg cattggtctt gaccaactct atcagagctt ggttgacggc 4320
aatttctgat atgcagcttg ggcgcagggt cgaatcgacc caatcgctcc atccggagcc 4380
gggactgtcg ggcgtacaca aatcgccgcg agaagcgccg ccgtctggac cgatggctgt 4440
gtagaagtag tcgccgatag tggaaaccga cgcccagca ctgctccgag ggcaaaggaa 4500
tgagtcgaga attcgctaga gggccctatt ctatagtgtc acctaattgc tagagctcgc 4560
tgatcagcct cgactgtgcc ttctagttag cagccatctg ttgtttgccc ctcccccggt 4620
ccttccttga ccttggaagg tggcactccc actgtccttt cctaataaaa tgaggaaatt 4680
gcactcgatt gtctgagtag gtgtcattct attctggggg gtgggggtgg gaggacagc 4740
aagggggagg atttggaaga caatagcagg catgctgagg gcccatttgc tgcagcgggc 4800
gcaataaaat atctttatct tcattacatc tgtgtgttgg ttttttgtgt gaatcgtaac 4860
taacatacgc tctccatcaa aacaaaacga aacaaaacaa actagcaaaa taggctgtcc 4920
ccagtgcagg tgcagggtgc agaacatttc tctatcgaag gatctgcgat cgctccgggt 4980
cccgtcagtg ggcagagcgc acatcgccca cagtcgccga gaagtggggg ggaggggtcg 5040
gcaattgaac cgtgtcctag agaaggtggc gcggggtaaa ctgggaaagt gatgctgtg 5100
actggctccg cctttttccc gaggttgggg gagaaccgta tataagtga gtagtcgccg 5160
tgaacgttct ttttcgcaac ggggtttgccc ccagaacaca gctgaagctt cgaggggtct 5220
gcactctccc cttacgcgcc cgccgcctta cctgagggcg ccatccacgc cgtttgagtc 5280
gcgttctgcc gcctcccgcc tgtggtgcc cctgaactgc gtccgccgtc taggtaagtt 5340
taaagctcag gtcgagaccg ggcctttgtc cggcgctccc ttgctcaact ttggagccta cctagactca 5400
gccggtcttc cagcctttgc ctgaccctgt ttgtcactac ctacgtcttt gtttcgtttg 5460
ctgttctgcg ccgttacaga tccaagctgt gaccggcgcc tacgtaagtg atatctacta 5520
gatttatcaa aaagagtgt gacttgtgag cgctcacaat tgatacttag attcatcgag 5580
agggacacgt cgactactaa cettcttctc tttcctacag ctgagat 5627

```

<210> 28
 <211> 553
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pMG plasmid from InvivoGen: EMCV IRES sequence

```

<400> 28
aacgttactg gccgaagccg cttggaataa ggccgggtgtg cgtttgtcta tatgttattt 60
tccaccatat tgccgtcttt tggcaatgtg agggcccgga aacctggccc tgtcttcttg 120
acgagcattc ctaggggtct tccccctctc gccaaaggaa tgcaaggctc gttgaatgtc 180
gtgaaggaa gcaatttgaac ggaagctctc tgaagacaaa caacgtctgt agcgaccctt 240
tgaggcagc ggaacccccc accctggcgac aggtgcctct gcggccaaaa gccacgtgta 300
taagatacac ctgcaaaggc ggcacaaccc cagtgcacag ttgtgagttg gatagttgtg 360
gaaagagtca aatggctctc ctcaagcgta ttcaacaagg ggctgaaggga tgcccagaag 420
gtacccattc gtaggggata tgatctgggg cctcggtgca catgctttac gtgtgtttag 480
tcgaggttaa aaaacgtcta gggcccccga accacgggga cgtggttttc ctttgaaaaa 540
cacgatgata ata 553

```

<210> 29
 <211> 4692
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pDSred1-N1 plasmid from Clontech

```

<400> 29
tagttattaa tagtaatcaa ttacgggggtc attagttcat agcccatata tggagttccg 60

```

-15-

cgttacataa	cttacggtaa	atggccccgc	tggetgaccg	cccaacgacc	cccgccatt	120
gacgtcaata	atgacgtatg	ttcccatagt	aacgcgaata	gggactttcc	attgacgtca	180
atgggtggag	tatttacggg	aaactgcccc	cttggcagta	catcaagtgt	atcatatgcc	240
aagtacgccc	cctattgacg	tcaatgacgg	taaattggccc	gcctgggcatt	atgcccagta	300
catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	tcgctattac	360
catggtgatg	cgggttttggc	agtacatcaa	tgggcgtgga	tagcgggttg	actcacgggg	420
atttccaagt	ctccaccccc	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	480
ggactttcca	aaatgtcgta	acaactccgc	cccatgacg	caaattgggcg	gtaggcggtg	540
acgggtgggag	gtctatataa	gcagagctgg	tttagtgaac	cgtcagatcc	gctagcgcta	600
ccggactcag	atctcgagct	caagcttcga	attctgcagt	cgacggtagc	gcgggcccg	660
gatccaccgg	tcgccaccat	ggtgcgctcc	tccaagaacg	tcatacaagg	gttcatgcgc	720
ttcaaggtgc	gcatggaggg	caccgtgaac	ggccacgag	tcgagatcga	gggcgaggcg	780
gagggccgcc	cctacgaggg	ccacaacacc	gtgaagctga	agggtgaccaa	gggcgggccc	840
ctgcccctcg	cctgggacat	cctgtccccc	cagttccagt	acgggtccaa	ggtgtacgtg	900
aagcaccggc	ccgacatccc	cgactacaag	aagctgtcct	tcgccgaggg	cttcaagttg	960
gagcgcgtga	tgaacttcga	ggacggcggc	gtggtgaccg	tgacccagga	ctcctccctg	1020
caggacggct	gcttcatcta	caaggtgaag	ttcatcggcg	tgaacttccc	ctccgacggc	1080
cccgtaatgc	agaagaagac	catgggctgg	gagggctcca	ccgagcgcc	gtaccccgcc	1140
gacggcgctg	tgaagggcga	gatccacaag	gcctgaagc	tgaaggacgg	cggccactac	1200
ctggtggagt	tcaagtccat	ctacatggcc	aagaagcccg	tgacgtgcc	cggctactac	1260
tacgtggact	ccaagctggg	ctcacctcc	cacaacgagg	actacaccat	cgtggagcga	1320
tacgagcgca	ccgagggccg	ccaccacctg	ttcctgtagc	ggccgcgact	ctagatcata	1380
atcagccata	ccacatttgt	agaggtttta	cttgctttta	aaaacctccc	acacctcccc	1440
ctgaacctga	aacataaaa	gaatgcaatt	gttgtgtgta	acttgtttat	tcagcgttat	1500
aatggttaca	aataaagcaa	tagcatcaca	aatttcacaa	ataaagcatt	tttttactg	1560
cattctagtt	gtgggtttgtc	caaactcatc	aatgtatctt	aaggcgtaaa	ttgtaaagcg	1620
taatattttg	ttaaaattcg	cgttaaaatt	ttgttaaata	agctcatttt	ttaaccaata	1680
ggccgaaatc	ggcaaaatcc	cttataaatc	aaaagaatag	accgagatag	ggttgagtgt	1740
tgtttccagtt	tggaaacaaga	gtccactatt	aaagaacgtg	gactccaacg	tcaaaaggcg	1800
aaaaaccgtc	tatcagggcg	atggcccact	acgtgaacca	tcacccta	caagtttttt	1860
ggggtcgagg	tgccgtaaag	cactaaatcg	gaaccctaaa	gggagccccc	gatttagagc	1920
ttgacgggga	aagccggcga	acgtggcgag	aaaggaagg	aagaaagcga	aaggagcggg	1980
cgtaagggcg	ctggcaagtg	tagcgggtcac	gctgtgctga	accaccacac	ccgcccgcgt	2040
taatgcgccc	ctacagggcg	cgtcaggtgg	cacttttcgg	ggaaatgtgc	gcggaacccc	2100
tatttgtttt	tttttctaaa	tacattcaaa	tatgtatccg	ctcatgagac	aataaccctg	2160
ataaatgtct	caataaatat	gaaaaaggaa	gagtcctgag	gcggaagaa	ccagctgtgg	2220
aatgtgtgtc	agttagggtg	tggaaagtcc	ccaggtctcc	cagcaggcag	aagtatgcaa	2280
agcatgcatc	tcaattagtc	agcaaccagg	tgtggaaagt	ccccaggctc	cccagcaggc	2340
agaagtatgc	aaagcatgca	tctcaattag	tcaggaacca	tagtcccgc	cctaattctc	2400
cccatcccg	ccctaactcc	gcccagttcc	gcccattctc	cgccccatgg	ctgactaatt	2460
ttttttat	atgcagaggg	cgaggccggc	tcggcctctg	agctattcca	gaagttagta	2520
ggaggctttt	tgggagggc	aggcttttgc	aaagatcgat	caagagacag	gatgaggatc	2580
gtttcgcatg	attgaacaag	atggatttgc	cgcaggttct	ccggccgctt	gggtggagag	2640
gctattcggc	tatgactggg	cacaacagac	aatcggtcgc	tctgatgccg	ccgtgttccg	2700
gctgtcagcg	caggggccc	cggttctttt	tgtcaagacc	gacctgtccg	gtgccctgaa	2760
tgaactgcaa	gacgagggc	cgcggtatc	gtggctggcc	acgacggggc	ttccttgccg	2820
agctgtgctc	gacgttgtca	ctgaagcggg	aagggaactg	ctgctattgg	gcgaagtgcc	2880
ggggcaggat	ctcctgtcat	ctcaccttgc	tctgcgcgag	aaagtatcca	tcattggctga	2940
tgcaatgcgg	cgggtgcata	cgcttgatcc	ggctacctgc	ccattcgacc	accaagcgaa	3000
acatcgcatc	gagcagcac	gtactcggat	ggaagccgg	cttgtcgatc	aggatgatct	3060
ggacgaagag	catcagggg	tcgcgccagc	cgaactgttc	gccaggctca	aggcgagcat	3120
gcccagcgcc	gaggatctcg	tcgtgaccca	tggcgatgac	tgcttgccga	atatcatggg	3180
ggaaaatggc	cgcttttctg	gattcatcga	ctgtggccgg	ctgggtgtgg	cggaccgcta	3240
tcaggacata	gcgttggtga	cccgtgatat	tgctgaagag	cttggcgggc	aatgggctga	3300
ccgcttcttc	gtgcttttac	gtatcgccgc	tcocgattcg	cagcgcacgc	ccttctatcg	3360
ccttcttgac	gagttcttct	gagcgggact	ctggggttcg	aaatgaccga	ccaagcgagc	3420
cccaacctgc	catcacgaga	tttcgattcc	accgcgcgct	tctatgaaag	gttgggcttc	3480
ggaaatcggt	tcggggacgc	cggctggatg	atcctccagc	gcggggatct	catgtgggag	3540
ttcttcgccc	accctagggg	gaggctaact	gaaacacgga	aggagacaat	accggaagga	3600
acccgcgcta	tgacggcaat	aaaaagacag	aataaaacgc	acggtgttgg	gtcgtttgtt	3660
cataaacggc	gggttcggtc	ccagggctgg	cactctgtcg	ataccccacc	gagaccccc	3720
tggggccaat	acgcccgcgt	ttcttccttt	tcocccaccc	accccccaag	ttcgggtgaa	3780
ggcccagggg	tcgcagccaa	cgtcggggcg	gcaggccctg	ccatagcctc	agggtactca	3840
tatacatggt	agattgattt	aaaacttcat	ttttatctta	aaaggatcta	ggtgaagatc	3900
ctttttgata	atctcatgac	caaaatccct	taacgtgagt	tttcgttcca	ctgagcgtca	3960
gaccccgtag	aaaagatcaa	aggatcttct	tgagatcctt	tttttctgcg	cgtaatctgc	4020
tgcttgcaaa	caaaaaaac	accgctacca	gcgggtgggtt	gtttgcccga	tcaagagcta	4080

-16-

ccaactcttt	ttccgaaggt	aactggcttc	agcagagcgc	agataccaaa	tactgtcctt	4140
ctagtgtagc	cgtagttagg	ccaccacttc	aagaactctg	tagcaccgcc	tacatacctc	4200
gctctgctaa	tcctgttacc	agtggctgct	gccagtggcg	ataagtctgt	tcttaccggg	4260
ttggactcaa	gacgatagtt	accggataag	gcgagcggt	cgggctgaac	gggggggtcg	4320
tgcacacagc	ccagcttgga	gcgaacgacc	tacaccgaac	tgagatacct	acagcgtgag	4380
ctatgagaaa	gcgccacgct	tcccgaaagg	agaaaggcgg	acaggtatcc	ggtaagcggc	4440
agggctcgaa	caggagagcg	cacgagggag	cttccagggg	gaaacgcctg	gtatctttat	4500
agtcctgtcg	ggtttcgcca	cctctgactt	gagcgtcgat	ttttgtgatg	ctcgtcaggg	4560
gggcggagcc	tatggaaaaa	cgccagcaac	gcggcctttt	tacggttcct	ggccttttgc	4620
tggccttttg	ctcacatgtt	ctttcctgcg	ttatcccctg	attctgtgga	taaccgtatt	4680
accgccatgc	at					4692

<210> 30

<211> 4257

<212> DNA

<213> Artificial Sequence

<220>

<223> pPur plasmid from Clontech

<400> 30

ctgtggaatg	tgtgtcagtt	aggggtgtgga	aagtccccag	gctccccagc	aggcagaagt	60
atgcaaagca	tgcattctcaa	ttagtcagca	accaggtgtg	gaaagtcccc	aggctcccca	120
gcaggcgagaa	gtatgcaaaag	catgcatctc	aattagtcag	caaccatagt	cccgccctta	180
actccgccca	tccccccctt	aactccgccc	agttccgccc	attctccgcc	ccatggctga	240
ctaatttttt	ttattttatgc	agaggccgag	gccgcctcgg	cctctgagct	attccagaag	300
tagtgaggag	gcttttttgg	aggcctaggc	ttttgcaaaa	agcttgcatg	cctgcaggtc	360
ggccgccacg	accggtgccc	ccaccatccc	ctgaccacag	ccctgacccc	ctcacaagga	420
gacgaccttc	catgaccgag	tacaagccca	cgggtgcgct	cgccacccgc	gacgacgtcc	480
cccgggcccgt	acgcacccct	gccgcgcgct	tcgcccacta	ccccgccacg	cgccacaccg	540
tcgacccgga	ccgccacatc	gagcgggtca	ccgagctgca	agaactcttc	ctcacgcgcg	600
tcgggctcga	catcggcaag	gtgtgggtcg	cggacgacgg	cgccgcgggt	gcggtctgga	660
ccacgcgcga	gagcgtcgaa	gcggggggcg	tggtcgccga	gatcggcccc	cgcattggccg	720
agttgagcgg	ttcccggctg	gccgcgcagc	aacagatgga	aggcctcctg	gcgccgcacc	780
ggcccaagga	gcccgcgtgg	ttcctggcca	ccgtcggcgt	ctcgcccgcac	caccagggga	840
agggctctgg	cagcgcgcgt	gtgctccccg	gagtgaggcg	ggccgagcgc	gccgggggtgc	900
ccgccttcct	ggagacctcc	gcgccccgca	acctccccct	ctacgagcgg	ctcggtttca	960
ccgtcaccgc	cgacgtcgag	gtgcccgaag	gaccgcgcac	ctggtgcatg	accgcgaagc	1020
ccggtgcctg	acgcccgcgc	cacgacccgc	agcgcgccac	cgaaaggagc	gcacgacccc	1080
atggctccga	ccgaagccga	cccgggcgcc	cccgccgacc	ccgcacccgc	ccccgaggcc	1140
caccgactct	agaggatcat	aatcagccat	accacatttg	tagaggtttt	acttgcttta	1200
aaaaacctcc	cacacctccc	cctgaacctg	aaacataaaa	tgaatgcaat	tggtgtgtgt	1260
aacttgttta	ttgcagctta	taatggttac	aaataaagca	atagcatcac	aaatttcaca	1320
aataaagcat	ttttttcact	gcattctagt	tgtggtttgt	ccaaactcat	caatgtatct	1380
tatcatgtct	ggatccccag	gaagctcctc	tgtgtcctca	taaaccctaa	cctcctctac	1440
ttgagaggac	attccaatca	taggctgccc	atccaccctc	tggtgcctcc	tggttaattag	1500
gtcacttaac	aaaaaggaaa	ttgggtaggg	gtttttcaca	gaccgctttc	taagggtaat	1560
tttaaaatat	ctgggaagtc	ccttccactg	ctgtgttcca	gaagtgttgg	taaacagccc	1620
acaaatgtca	acagcagaaa	catacaagct	gtcagctttg	cacaagggcc	caacaccctg	1680
ctcatcaaga	agcactgtgg	ttgctgtgtt	agtaatgtgc	aaaacaggag	gcacattttc	1740
cccacctgtg	taggttccaa	aatatctagt	gttttcattt	ttacttggat	caggaaccca	1800
gcactccact	ggataagcat	tatccttatc	caaaacagcc	ttgtggtcag	tggtcatctg	1860
ctgactgtca	actgtagcat	tttttggggg	tacagtttga	gcaggatatt	tggtcctgta	1920
gtttgtctaa	acaccttgca	gctccaaagg	ttccccacca	acagcaaaaa	aatgaaaatt	1980
tgaccttgta	atgggttttc	cagcaccatt	ttcatgagtt	ttttgtgtcc	ctgaatgcaa	2040
gtttaacata	gcagttaccc	caataacctc	agttttaaca	gtaacagctt	cccacatcaa	2100
aatatttcca	caggttaagt	cctcatttaa	attaggcaaa	ggaattcttg	aagacgaaag	2160
ggcctcgtga	tacgcctatt	tttataggtt	aatgtctaga	taataatggg	ttcttagacg	2220
tcagggtggca	cttttcgggg	aaatgtgcgc	ggaaccctta	tttgtttatt	tttctaaata	2280
cattcaataa	tgatccgct	catgagacaa	tacccttgat	aaatgcttca	ataaatattga	2340
aaaagggaaga	gtatagttat	tcaacatttc	cgtgtcgccc	ttattccctt	ttttcgggca	2400
ttttgccttc	ctgtttttgc	tcacccagaa	acgctggtga	aagtaaaaaga	tgctgaagat	2460
cagttgggtg	ccgagtggtg	ttacatcgaa	ctggatctca	acagcggtaa	gatccttgag	2520
agttttcgcc	ccgaagaacg	ttttccaatg	atgagcattc	ttaaagtctc	gctatgtggc	2580
gcggtattat	cccgtgttga	cgccggggcaa	gagcaactcg	gtcgccgcac	acactattct	2640
cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	tggtcatgaca	2700
gtaagagaat	tatgcagtcg	tgccataaac	atgagtgata	acactgcggc	caacttactt	2760

-17-

ctgacaacga	tccgaggacc	gaaggagcta	accgcttttt	tgcacaacat	gggggatcat	2820
gtaactcgcc	ttgatcggtg	ggaaccggag	ctgaatgaag	ccataccaaa	cgacgagcgt	2880
gacaccacga	tgcctgcagc	aatggcaaca	acgttgcgca	aactattaac	tggcggaacta	2940
cttactctag	cttcccgcca	acaattaata	gactggatgg	aggcggataa	agttgcagga	3000
ccacttctgc	gctcggccct	tccggtctggc	tgggtttattg	ctgataaatc	tggagccggt	3060
gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	atggtaagcc	ctcccgtatc	3120
gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	acagatcgct	3180
gagataggtg	ctcactgat	taagcattgg	taactgtcag	accaagttta	ctcatatata	3240
cttttagattg	atttaaaaact	tcattttttaa	tttaaaaagga	tctaggtgaa	gacccctttt	3300
gataatctca	tgaccaaaaa	cccttaacgt	gagttttcgt	tccactgagc	gtcagacccc	3360
gtagaaaaga	tcaaaggatc	ttcttgagat	cttttttttc	tgcgcgtaat	ctgctgcttg	3420
caaacaaaaa	aaccaccgct	accagcgggtg	gtttgttttgc	cggatcaaga	gctaccaact	3480
ctttttccga	aggtaactgg	cttcagcaga	gcgcagatac	caaatactgt	ccttctagt	3540
tagccgtagt	taggcaccca	cttcaagaac	tctgtagcac	cgctacata	cctcgctctg	3600
ctaactctgt	taccagtggc	tgtgccagt	ggcgataagt	cgtgtcttac	cgggttggtac	3660
tcaagacgat	agttaccgga	taaggcgcag	cggctcgggt	gaacgggggg	tctgtgcaca	3720
cagcccagct	tggagcgaac	gaactacacc	gaactgagat	acctacagcg	tgagctatga	3780
gaaagcgcca	cgcttcccga	agggagaaag	gcggacaggt	atccggtaag	cggcaggggtc	3840
ggaacaggag	agcgcacgag	ggagcttcca	gggggaaacg	cctgggtatct	ttatagtcct	3900
gtcgggtttc	tccacctctg	acttgagcgt	cgatttttgc	gatgctcgtc	agggggggcgg	3960
agcctatgga	aaaacgccag	caacgcggcc	tttttacggt	tccctggcctt	ttgctggcct	4020
tttgcctaca	tgttctttcc	tgcgttatcc	cctgattctg	tggataaccg	tattaccgcc	4080
tttgagtgag	ctgataccgc	tgcgcgcagc	cgaacgaccg	agcgcagcga	gtcagtgagc	4140
gaggaagcgg	aagagcgcc	gatgcgggtat	tttctcctta	cgcactctgtg	cgggtatttca	4200
caccgcata	ggtgcactct	cagtacaatc	tgtctctgatg	cgcgatagtt	aagccag	4257

<210> 31

<211> 8136

<212> DNA

<213> Artificial Sequence

<220>

<223> pWE15 cosmid vector

<300>

<308> GenBank X65279

<309> 1995-04-14

<400> 31

ctatagttag	tcgattattg	cggccgcgaa	ttcttgaaga	cgaaggggcc	tcgtagatag	60
cctattttta	taggttaaat	tcattgataa	aatgggtttc	tagacgtcag	gtggcacttt	120
tccgggaaat	gtgcgcggaa	ccctattttg	tttatttttc	taaatacatt	caaataatga	180
tccgctcatg	agacaataac	cctgataaat	gcttcaataa	tattgaaaaa	ggaagagtat	240
gagtattcaa	catttccgtg	tccctttttt	cccttttttt	gcggcatttt	gcttccctgt	300
tttgcctacc	cagaaaacgt	ggtgaaagta	aaagatcgct	aagatcagtt	gggtgcacga	360
gtgggttaca	tcgaactgga	tctcaacagc	ggtaagatcc	ttgagagttt	tcccccga	420
gaacgttttc	caatgatgag	cactttttaa	gttctgctat	gtggcgcggt	attatcccg	480
gttgacgcgc	ggcaagagca	actcggtcgc	cgcatacact	attctcagaa	tgacttggtt	540
gagtactcac	cagtcacaga	aaagcatctt	acggatggca	tgacagtaag	agaattatgc	600
agtgcctgca	taaccatgag	tgataaacct	gcggccaact	tacttctgac	aacgatcgga	660
ggaccgaagg	agctaaccgc	ttttttgcac	aacatggggg	atcatgtaac	tgccttgat	720
cgttgggaac	cggagctgaa	tgaagccata	ccaaacgacg	agcgtgacac	cacgatgcct	780
gcagcaatgg	caacaacgtt	gcgcaaaact	ttaactggcg	aactacttac	tctagcttcc	840
cggcaacaat	taatagactg	gatggaggcg	gataaagtgt	caggaccact	tctgcgctcg	900
gcccttccgg	ctggctgggt	tattgctgat	aaatctggag	ccggtgagcg	tgggtctcgc	960
ggtatcattg	cagcactggg	gccagatggg	aagccctccc	gtatcgtagt	tatctacacg	1020
acggggagtc	aggcaactat	ggatgaacga	aatagacaga	tgcgtgagat	aggtgcctca	1080
ctgattaagc	atttggaact	gtcagaccaa	gtttactcat	atatacttta	gattgattta	1140
aaaacttcatt	tttaatttaa	aaggatctag	gtgaagatcc	tttttgataa	tctcatgacc	1200
aaaatccctt	aacgtgagtt	ttcggtccac	tgagcgtcag	accccgtaga	aaagatcaaa	1260
ggatcttctt	gagatccctt	ttttctgcgc	gtaatctgct	gcttgcaaac	aaaaaaacca	1320
ccgctaccag	cgggtggtttg	tttgccggat	caagagctac	caactctttt	tccgaaggta	1380
actgcttcca	gcagagcgca	gataccaaat	actgtccttc	tagtgtagcc	gtagttagcc	1440
caccacttca	agaactctgt	agcaccgcct	acatacctcg	ctctgcta	cctgttacca	1500
gtggctgctg	ccagtggcga	taagtcgtgt	cttaccgggt	tggactcaag	acgatagtta	1560
cgggataaag	cgcagcggtc	gggctgaacg	gggggttcgt	gcacacagcc	cagcttggag	1620
cgaacgacct	acaccgaact	gagataccta	cagcgtgagc	tatgagaaag	cggccagcctt	1680

-18-

ccgaagggag	aaagggcgac	aggtatccgg	taagcggcag	ggcgggaaca	ggagagcgca	1740
cgagggagct	tccaggggga	aacgcctggg	atcttttatag	tcctgtcggg	gtttcgccac	1800
ctctgacttg	agcgtcgatt	tttgtgatgc	tcgtcagggg	ggcggagcct	atggaaaaac	1860
gccagcaacg	cggccttttt	acggttcctg	gcctttttgt	ggccttttgc	tcacatgttc	1920
tttctcgctg	tatccccgta	ttctgtggat	aaccgtatta	cgccttttga	gtgagctgat	1980
accgctcgcc	gcagccgaac	gaccgagcgc	agcaggtcag	tgagcgagga	agcgggaagag	2040
cgctgacttc	cgcgtttcca	gacttttacga	aacacggaaa	ccgaagacca	ttcatgttgt	2100
tgctcagggt	gcagacgttt	tgacgcagca	gtcgtttcac	gttcgctcgc	gtatcgggtga	2160
ttcattctgc	taaccagtaa	ggcaaccccg	ccagcctagc	cgggtcctca	acgacaggag	2220
cacgatcatg	cgcacccgtc	agatccagac	atgataagat	acattgatga	gtttgggacaa	2280
accacaacta	gaatgcagtg	aaaaaaatgc	tttatttgtg	aaatttgtga	tccttattta	2340
ttatttgtta	ccattataag	ctgcaataaa	caagttaaca	acaacaattg	cattcatttt	2400
atgtttcagg	ttcaggggga	gggtgtgggag	gtttttttaa	gcaagtaaaa	cctctacaaa	2460
tggtgtatgg	ctgattatga	tctctagtca	aggcatttga	catcaaata	tccttattta	2520
cccctttaca	aattaaaaag	ctaaagggtac	acaatttttg	agcatagtta	ttaatagcag	2580
acactctatg	cctgtgtgga	gtaagaaaaa	acagtattgt	atgattataa	ctgttatgcc	2640
tacttataaa	ggttacagaa	tatttttcca	taattttctt	gtatagcagt	gcagcttttt	2700
cctttgtggg	gtaaatagca	aagcaagcaa	gagttctatt	actaaacaca	gcatgactca	2760
aaaaacttag	caattctgaa	ggaaagtcct	tggggtcttc	tacctttctc	ttcttttttg	2820
gaggagttag	atgttgagag	tcagcagtag	cctcatcctc	actagatggc	atttctctcg	2880
agcaaaacag	gttttcctca	ttaaaggcat	tccaccactg	ctcccatcca	tcagttccat	2940
agggttggaat	ctaaaataca	caaacaatta	gaatcagtag	tttaacacat	tatacactta	3000
aaaattttat	atttacctta	gagcttttaa	tctctgtagg	tagtttgtcc	aattatgtca	3060
caccacagaa	gtaagggttc	ttcacaaaag	tccggaccaa	agcggccatc	gtgctctccc	3120
actcctgcag	ttcgggggca	tggatgcgcg	gatagccgct	gctgggttcc	tggatgcoga	3180
cggatttgca	ctgcgggtag	aactcgcgag	gtcgtccagc	ctcaggcagc	agctgaacca	3240
actcgcgagg	ggatcgagcc	cgggggtggc	gaagaactcc	agcatgagat	ccccgcgctg	3300
gaggatcatc	cagccggcgt	cccggaaaac	gattccgaag	cccaaccttt	catagaaggc	3360
ggcgggtgaa	tcgaaatctc	gtgatggcag	gttgggctgc	gcttggtcgg	tcatttcgaa	3420
cccagagtc	ccgctcagaa	gaactcgta	agaaggcgat	agaaggcgat	gcgctgcgaa	3480
tcgggagcgg	cgataccgta	aagcacgagg	aagcggtcag	cccattcgcc	gccaagctct	3540
tcagcaatat	cacgggtagc	caacgctatg	tcctgatagc	ggtcggccac	acccagccgg	3600
ccacagtcga	tgaatccaga	aaagcggcca	ttttccacca	tgatattcgg	caagcaggca	3660
tcgccatggg	tcacgacgag	atcctcgccg	tcgggatgcg	cgccttgagc	ctggcgaaaca	3720
gttcggctgg	cgcgagcccc	tgatgctctt	cgtccagatc	atcctgatcg	acaagaccgg	3780
cttccatccg	agtcagtgct	cgctcgatgc	gatgtttcgc	ttggtggctg	aatggcgctg	3840
tagccggatc	aagcgtatgc	agccgcccga	ttgcatcagc	catgatggat	actttctcgg	3900
caggagcaag	gtgagatgac	aggagatcct	gccccggcac	ttcgcccaat	agcagccagt	3960
cccttcccgc	ttcagtgaca	acgtcgagca	cagctgcgca	aggaacgccc	gtcgtggcca	4020
gccacgatag	ccgcgtgcc	tcgtcctgca	gttcattcag	ggcaccggac	aggctcggct	4080
tgacaaaaag	aaccgggcgc	ccctgcgtcg	acagccggaa	cacggcggca	tcagagcage	4140
cgattgtctg	ttgtgccag	tcatagcoga	atagcctctc	cacccaagcg	gccggagaa	4200
ctgcgtgcaa	tccatcttgt	tcaatcatgc	gaaacgatcc	tcattcctgtc	tcttgatcag	4260
atcttgatcc	cctgcgccat	cagatccttg	gcccgaagaa	agccatccag	tttactttgc	4320
agggcttccc	aaccttacca	gagggcgccc	cagctggcaa	ttccggttcg	cttgctgttc	4380
ataaaaccgc	ccagtctagc	tatcgccatg	taagccact	gcaagctacc	tgctttctct	4440
ttgcgcttgc	gttttccctt	gtccagatag	cccagtagct	gacattcatc	cggggctcagc	4500
accgtttctg	cggactggct	ttctacgtgt	tcgcttctc	ttagcagccc	ttgcgcctg	4560
agtgtttgcg	gcagcgtgaa	agctttttgc	aaaagcctag	gcctccaaaa	aagcctcctc	4620
actacttctg	gaatagctca	gaggccgagg	cggcctaaat	aaaaaaaaat	agtcagccat	4680
ggggcgagga	atgggcggaa	ctgggcggag	ttagggcgcg	gatgggcgga	gttaggggag	4740
ggactatggg	tgctgactaa	ttgagatgca	tgctttgcat	acttctgcct	gtcggggagc	4800
ctggggactt	tccacacctg	gttgctgact	aattgagatg	catgctttgc	atacttctgc	4860
ctgctgggga	gcctggggac	tttccacacc	ctaactgaca	cacattccac	agccgatct	4920
gcaggaccca	acgctgcccg	agatgcgcgc	cgtgcggctg	ctggagatgg	cggacgcgat	4980
ggatatgttc	tgccaagggt	tggtttgcgc	attcacagtt	ctccgcaaga	attgatggc	5040
tccaattctt	ggagtgggtg	atccgttagc	gaggtgccgc	cggcttccat	tcaggtcgag	5100
gtggcccgcc	tccatgcacc	gcgacgcaac	gcggggaggc	agacaaggta	tagggcgccg	5160
cctacaatcc	atgccaaccc	gttccatgtg	ctcgccgagg	cgcataaatc	gccgtgacga	5220
tcagcggtcc	aatgatcgaa	gttaggtggg	taagagccgc	gagcgatcct	tgaagctgtc	5280
cctgatgggt	gtcatctacc	tgccctggaca	gcattggcctg	caacgcggca	tcccgatgcc	5340
gcgggaagcg	agaagaatca	taattggggaa	ggccatccag	cctcgcgctg	cgaacgccag	5400
caagacgtag	ccagcgcgt	cgggcgcgca	tgccggcgat	aatggcctgc	ttctcgccga	5460
aacgtttggg	ggcggggacca	gtgacgaagg	cttgagcgag	ggcgtgcaag	attccgcaata	5520
ccgcaagcga	caggccgatc	atcgctcgcg	tccagcgaaa	gcggtcctcg	ccgaaaatga	5580
cccagagcgc	tgccggcacc	gttcctacga	gttgcatgat	aaagaagaca	gtcataagtg	5640
cgccgacgat	agtcattgcc	cgcgcccacc	ggaaggagct	gactgggttg	aaggctctca	5700

-19-

agggcatcgg	tgcagcgtct	cccttatgcg	actcctgcac	taggaagcag	cccagtagta	5760
gggtgagggc	gttgagcacc	gccgcccga	ggaatgggtg	atgcaaggag	atggcgccca	5820
acagtccccc	ggccacgggc	ctgccacgat	acccacgccc	aaacaagcgc	tcagagccc	5880
gaagtggcga	gcccgatctt	ccccatcgg	gatgtcggcg	atatagggcg	cagcaaccgc	5940
acctgtggcg	ccggtgatgc	cggccacgat	gcgtccggcg	tagaggatct	tggcagtcac	6000
agcatgcgca	tatccatgct	tgcacacgat	gctcacaaag	taggtgaatg	cgcaatgtag	6060
taccacacat	gtcatcgctt	tccactgctc	tcgcgaataa	agatggaaaa	tcaatctcat	6120
ggtaatatgc	catgaaaaatc	cttgtattca	taaatcctcc	aggtagctat	atgcaaattg	6180
aaacaaaaga	gatgggtgatc	tttctaagag	atgatggaat	ctcccttcag	tatcccgatg	6240
gtcaatgcgc	tggatatggg	atagatggga	atatgtgat	ttttatggga	cagagttgcg	6300
aactgttccc	aactaaaaatc	attttgcacg	atcagcgcac	tacgaacttt	acccacaaat	6360
agtcaggtaa	tgaatcctga	tataaagaca	gggtgataaa	tcagtcttct	acgcgcacgc	6420
cacgcgcaca	ccgtagaaaag	tctttcagtt	gtgagcctgg	gcaaaccggt	aactttcggc	6480
ggctttgtcg	tcacgtctaa	tcacgtctaa	aaggaaataa	atcatgggtc	ataaaattat	6540
cacgttgtcc	ggcgcggcga	cggatgttct	gtatgcgctg	tttttccgtg	gcgcgttgct	6600
gtctgggtgat	ctgccttcta	aatctggcac	agccgaattg	cgcgagcttg	gttttgcgtg	6660
aaccagacac	acagcaactg	aataccagaa	agaaaatcac	tttacctttc	tgacatcaga	6720
agggcagaaa	tttgccgttg	aacacctggt	caatacgcgt	tttggtgagc	agcaatattg	6780
cgcttcgatg	acgcttgccg	ttgagattga	tacctctgct	gcacaaaagg	caatcgacga	6840
gctggaccag	cgcattcgtg	acaccgtctc	cttcgaactt	attcgcaatg	gagtgctcatt	6900
catcaaggac	gccgtatctg	caaatgggtg	tatccacgca	gcggcaatcg	aaacacctca	6960
gcccggtgacc	aatatctaca	acatcagcct	tggtatccag	cgtgatgagc	cagcgcagaa	7020
caaggttaacc	gtcagtgccg	ataagttcaa	agttaaacct	ggtgttgata	ccaacattga	7080
aacgttgtatc	gaaaacgcgc	tgaaaaacgc	tgctgaatgt	gcggcgctgg	atgtcacaaa	7140
gcaaatggca	gcagacaaga	aagcgatgga	tgaactggct	tcctatgtcc	gcacggccat	7200
catgatggaa	tggttccccg	gtggtgttat	ctggcagcag	tgccgtcgat	agtatgcaat	7260
tgataattat	tatcatttgc	gggtcctttc	cggcgatccg	ccttgttacg	ggcgccgcac	7320
ctcgcgggtt	ttcgctattt	atgaaaattt	tccggtttta	ggcgtttccg	ttcttcttcg	7380
tcataaactta	atgtttttat	ttaaaatacc	ctctgaaaag	aaaggaaaacg	acaggtgctg	7440
aaagcgagct	ttttggcctc	tgctgtttcc	tttctctgtt	tttgctccgtg	gaatgaacaa	7500
tggaagtcaa	caaaaagcag	ctggctgaca	ttttcgggtg	gagtatccgt	accattcaga	7560
actggcagga	acagggaatg	cccgttctgc	gaggcggtgg	caagggtaat	gaggtgcttt	7620
atgactctgc	cgcgctcata	aaatggtatg	ccgaaaaggga	tgctgaaatt	gagaacgaaa	7680
agctgcgccc	ggaggttgaa	gaactgcggc	aggccagcga	ggcagatcca	caggacgggt	7740
gtggtcgcca	tgatcgcgta	gtcgatagtg	gtccaaagta	gcgaagcgag	caggactggg	7800
cggcggcaaa	gcggtcggac	agtgtccga	gaacgggtgc	gcatagaaat	tgcatcaacg	7860
catatagcgc	tagcagcacg	ccatagtgc	tggcgatgct	gtcggaaatgg	acgatatccc	7920
gcaagaggcc	cggcagtagc	ggcataacca	agcctatgcc	tacagcatcc	agggtgacgg	7980
tgccgagagc	gacgtgagc	gcattgttag	atctcataca	cgggtgcctga	ctgcgttagc	8040
aatttaactg	tgataaaacta	ccgcattaaa	gcttatcgat	gataagcggg	caaacatgag	8100
aattcgcggc	cgcaattaac	cctcactaaa	ggtatcc			8136

<210> 32
 <211> 2713
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pNEB193 plasmid

<400> 32						
tcgcgcgttt	cggatgatgac	ggtgaaaacc	tctgacacat	gcagctcccg	gagacgggtca	60
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	120
ttggcgggtg	tcggggcgtg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	180
accatatgcg	gtgtgaaata	ccgcacagat	cggttaaggag	aaaataccgc	atcaggcgcc	240
attcgccatt	caggctgcgc	aactgttggg	aagggcgatc	ggtgcggggc	tcttcgctat	300
tacgccagct	ggcgaaggag	ggatgtgctg	caaggcgatt	aagttgggtg	acgccagggt	360
tttcccagtc	acgacgttgt	aaaacgacgc	ccagtgaatt	cgagctcggt	acccgggggc	420
gcgcgggatc	cttaatttaag	tctagagtcg	actgtttaaa	cctgcaggca	tgcaagcttg	480
gcgtaatcat	ggtcatagct	gtttcctgtg	tgaattgtgt	atccgctcac	aattccacac	540
aacatacgag	ccggaagcat	aaagtgtaaa	gcctgggggtg	cctaattgagt	gagctaactc	600
acattaattg	cgttcgcgtc	actgcccgcg	ttccagtcgg	gaaacctgtc	gtgccaagctg	660
cattaatgaa	tcggccaacg	cgcggggaga	ggcggtttgc	gtattgggcg	ctcttcgctg	720
tcctcgctca	ctgactcgct	gcgctcggtc	gttcgggtgc	ggcgagcggg	atcagctcac	780
tcaaaggcgg	taatacgggt	atccacagaa	tcaggggata	acgcaggaaa	gaacatgtga	840
gcaaaaggcc	agcaaaaggc	caggaaccgt	aaaaaggccg	cgttgctggc	gtttttccat	900
aggctccgcc	cccctgacga	gcatacacia	aatcgacgct	caagtcagag	gtggcgaaac	960

-20-

ccgacaggac	tataaagata	ccaggcgttt	ccccctggaa	gctccctcgt	gcgctctcct	1020
gttccgaccc	tgccgcttac	cggatacctg	tccgcctttc	tcccttcggg	aagcgtggcg	1080
ctttctcata	gctcacgctg	taggtatctc	agttcgggtg	aggctcgttcg	ctccaagctg	1140
ggctgtgtgc	acgaaccccc	cgttcagccc	gaccgctgcg	ccttatccgg	taactatcgt	1200
cttgagtcca	acccggtaag	acacgactta	tcgccactgg	cagcagccac	tggtaacagg	1260
attagcagag	cgagggtatgt	aggcgggtgct	acagagttct	tgaagtgggtg	gcctaactac	1320
ggctacacta	gaaggacagt	atttggtatc	tgcgctctgc	tgaagccagt	taccttcgga	1380
aaaagagttg	gtagctcttg	atccggcaaaa	caaaccaccg	ctggtagcgg	tggttttttt	1440
gtttgcaagc	agcagattac	gcgcagaaaa	aaaggatctc	aagaagatcc	tttgatcttt	1500
tctacggggg	ctgacgctca	gtggaacgaa	aactcacgtt	aagggatttt	gggtcatgaga	1560
ttatcaaaaa	ggatcttcac	ctagatcctt	ttaaattaaa	aatgaagttt	taaatcaatc	1620
taaagtatat	atgagtaaac	ttggtctgac	agttaccaat	gcttaatcag	tgaggcacct	1680
atctcagcga	tctgtctatt	tcgttcatcc	atagttgcct	gactccccgt	cgtgtagata	1740
actacgatac	gggagggcct	accatctggc	cccagtgcgtg	caatgatacc	gcgagacca	1800
cgtccaccgg	ctccagattt	atcagcaata	aaccagccag	ccggaagggc	cgagcgcaga	1860
agtggtcctg	caactttatc	cgcctccatc	cagctcatta	attgttgccg	ggaagctaga	1920
gtaagtattg	cgccagttta	tagtttgccg	aacgttggtg	ccattgctac	aggcatcgtg	1980
gtgtcacgct	cgtcgtttgg	tatggcttca	ttcagctccg	gttcccaacg	atcaaggcga	2040
gttacatgat	cccccatggt	gtgcaaaaaa	gcggttagct	ccttcgggtcc	tccgatcggt	2100
gtcagaagta	agttggccgc	agtgttatca	ctcatgggta	tggcagcact	gcataattct	2160
cttactgtca	tgccatccgt	aagatgcctt	tctgtgactg	gtgagtactc	aaccaagtca	2220
ttctgagaat	agtgtatgcg	gcgaccgagt	tgctcttgcc	cggcgtcaat	acggggataat	2280
accgcgccac	atagcagaac	tttaaaagtg	ctcatcattg	gaaaacgttc	ttcggggcga	2340
aaactctcaa	ggatctttacc	gctgttgaga	tccagttcga	tgtaaccac	tcgtgcaccc	2400
aactgatctt	cagcatcttt	tactttcacc	agcgtttctg	ggtgagcaaa	aacaggaagg	2460
caaaatgccg	caaaaaaggg	aataagggcg	acacggaaat	ggtgaatact	catactcttc	2520
ctttttcaat	attattgaag	catttatcag	ggttattgtc	tcatgagcgg	atacatattt	2580
gaatgtatct	agaaaaataa	acaaataggg	gttccgcgca	catttccccg	aaaagtgcc	2640
cctgacgtct	aagaaaccat	tattatcatg	acattaacct	ataaaaaatg	gcgtatcacg	2700
aggccctttc	gtc					2713

<210> 33
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> attP

<400> 33
 cagctttttt atactaagtt g 21

<210> 34
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> attB

<400> 34
 ctgctttttt atactaactt g 21

<210> 35
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> attL

<400> 35
 ctgctttttt atactaagtt g 21

<210> 36
 <211> 21
 <212> DNA

-21-

<213> Artificial Sequence

<220>

<223> attR

<400> 36

cagctttttt atactaactt g

21

<210> 37

<211> 1071

<212> DNA

<213> Artificial Sequence

<220>

<223> Integrase E174R

<221> CDS

<222> (1)...(1071)

<223> Nucleotide sequence encoding Integrase E147R

<400> 37

atg gga aga agg cga agt cat gag cgc cgg gat tta ccc cct aac ctt	48
Met Gly Arg Arg Asn Ser His Glu Arg Arg Asp Leu Pro Pro Asn Leu	
1 5 10 15	
tat ata aga aac aat gga tat tac tgc tac agg gac cca agg acg ggt	96
Tyr Ile Arg Asn Gly Tyr Tyr Cys Tyr Arg Asp Pro Arg Thr Gly	
20 25 30	
aaa gag ttt gga tta ggc aga gac agg cga atc gca atc act gaa gct	144
Lys Glu Phe Gly Leu Gly Arg Asp Arg Arg Ile Ala Ile Thr Glu Ala	
35 40 45	
ata cag gcc aac att gag tta ttt tca gga cac aaa cac aag cct ctg	192
Ile Gln Ala Asn Ile Glu Leu Phe Ser Gly His Lys His Lys Pro Leu	
50 55 60	
aca gcg aga atc aac agt gat aat tcc gtt acg tta cat tca tgg ctt	240
Thr Ala Arg Ile Asn Ser Asp Asn Ser Val Thr Leu His Ser Trp Leu	
65 70 75 80	
gat cgc tac gaa aaa atc ctg gcc agc aga gga atc aag cag aag aca	288
Asp Arg Tyr Glu Lys Ile Leu Ala Ser Arg Gly Ile Lys Gln Lys Thr	
85 90 95	
ctc ata aat tac atg agc aaa att aaa gca ata agg agg ggt ctg cct	336
Leu Ile Asn Tyr Met Ser Lys Ile Lys Ala Ile Arg Arg Gly Leu Pro	
100 105 110	
gat gct cca ctt gaa gac atc acc aca aaa gaa att gcg gca atg ctc	384
Asp Ala Pro Leu Glu Asp Ile Thr Lys Lys Glu Ile Ala Ala Met Leu	
115 120 125	
aat gga tac ata gac gag ggc aag gcg gcg tca gcc aag tta atc aga	432
Asn Gly Tyr Ile Asp Glu Gly Lys Ala Ala Ser Ala Lys Leu Ile Arg	
130 135 140	
tca aca ctg agc gat gca ttc cga gag gca ata gct gaa ggc cat ata	480
Ser Thr Leu Ser Asp Ala Phe Arg Glu Ala Ile Ala Glu Gly His Ile	
145 150 155 160	
aca aca aac cat gtc gct gcc act cgc gca gca aaa tct aga gta agg	528
Thr Thr Asn His Val Ala Ala Thr Arg Ala Ala Lys Ser Arg Val Arg	
165 170 175	
aga tca aga ctt acg gct gac gaa tac ctg aaa att tat caa gca gca	576
Arg Ser Arg Leu Thr Ala Asp Glu Tyr Leu Lys Ile Tyr Gln Ala Ala	

-22-

180										185										190										
gaa	tca	tca	cca	tgt	tgg	ctc	aga	ctt	gca	atg	gaa	ctg	gct	gtt	gtt	624														
Glu	Ser	Ser	Pro	Cys	Trp	Leu	Arg	Leu	Ala	Met	Glu	Leu	Ala	Val	Val															
		195					200					205																		
acc	ggg	caa	cga	gtt	ggg	gat	tta	tgc	gaa	atg	aag	tgg	tct	gat	atc	672														
Thr	Gly	Gln	Arg	Val	Gly	Asp	Leu	Cys	Glu	Met	Lys	Trp	Ser	Asp	Ile															
	210					215					220																			
gta	gat	gga	tat	ctt	tat	gtc	gag	caa	agc	aaa	aca	ggc	gta	aaa	att	720														
Val	Asp	Gly	Tyr	Leu	Tyr	Val	Glu	Gln	Ser	Lys	Thr	Gly	Val	Lys	Ile															
	225				230					235					240															
gcc	atc	cca	aca	gca	ttg	cat	att	gat	gct	ctc	gga	ata	tca	atg	aag	768														
Ala	Ile	Pro	Thr	Ala	Leu	His	Ile	Asp	Ala	Leu	Gly	Ile	Ser	Met	Lys															
				245					250					255																
gaa	aca	ctt	gat	aaa	tgc	aaa	gag	att	ctt	ggc	gga	gaa	acc	ata	att	816														
Glu	Thr	Leu	Asp	Lys	Cys	Lys	Glu	Ile	Leu	Gly	Gly	Glu	Thr	Ile	Ile															
			260					265					270																	
gca	tct	act	cgt	cgc	gaa	ccg	ctt	tca	tcc	ggc	aca	gta	tca	agg	tat	864														
Ala	Ser	Thr	Arg	Arg	Glu	Pro	Leu	Ser	Ser	Gly	Thr	Val	Ser	Arg	Tyr															
		275					280					285																		
ttt	atg	cgc	gca	cga	aaa	gca	tca	ggg	ctt	tcc	ttc	gaa	ggg	gat	ccg	912														
Phe	Met	Arg	Ala	Arg	Lys	Ala	Ser	Gly	Leu	Ser	Phe	Glu	Gly	Asp	Pro															
	290					295					300																			
cct	acc	ttt	cac	gag	ttg	cgc	agt	ttg	tct	gca	aga	ctc	tat	gag	aag	960														
Pro	Thr	Phe	His	Glu	Leu	Arg	Ser	Leu	Ser	Ala	Arg	Leu	Tyr	Glu	Lys															
	305				310					315					320															
cag	ata	agc	gat	aag	ttt	gct	caa	cat	ctt	ctc	ggg	cat	aag	tcg	gac	1008														
Gln	Ile	Ser	Asp	Lys	Phe	Ala	Gln	His	Leu	Leu	Gly	His	Lys	Ser	Asp															
				325					330					335																
acc	atg	gca	tca	cag	tat	cgt	gat	gac	aga	ggc	agg	gag	tgg	gac	aaa	1056														
Thr	Met	Ala	Ser	Gln	Tyr	Arg	Asp	Asp	Arg	Gly	Arg	Glu	Trp	Asp	Lys															
			340					345					350																	
att	gaa	atc	aaa	taa												1071														
Ile	Glu	Ile	Lys	*																										
			355																											

<210> 38

<211> 356

<212> PRT

<213> Artificial Sequence

<220>

<223> Integrase E147R

<400> 38

Met	Gly	Arg	Arg	Arg	Ser	His	Glu	Arg	Arg	Asp	Leu	Pro	Pro	Asn	Leu
1				5					10					15	
Tyr	Ile	Arg	Asn	Asn	Gly	Tyr	Tyr	Cys	Tyr	Arg	Asp	Pro	Arg	Thr	Gly
			20					25					30		
Lys	Glu	Phe	Gly	Leu	Gly	Arg	Asp	Arg	Arg	Ile	Ala	Ile	Thr	Glu	Ala
		35					40				45				
Ile	Gln	Ala	Asn	Ile	Glu	Leu	Phe	Ser	Gly	His	Lys	His	Lys	Pro	Leu
	50					55				60					
Thr	Ala	Arg	Ile	Asn	Ser	Asp	Asn	Ser	Val	Thr	Leu	His	Ser	Trp	Leu
	65				70				75					80	
Asp	Arg	Tyr	Glu	Lys	Ile	Leu	Ala	Ser	Arg	Gly	Ile	Lys	Gln	Lys	Thr

[illegible]

```
<220>
<221> CDS
<222> (45)...(737)
<223> Nucleotide sequence encoding red fluorescent
      protein (FP593)
```

<400> 39																
agttttcagcc agtgacaggg tgagctgccca ggtattcttaa caag atg agt tgt tcc																56
Met Ser Cys Ser																
1																
aag aat gtg atc aag gag ttc atg agg ttc aag gtt cgt atg gaa gga																104
Lys Asn Val Ile Lys Glu Phe Met Arg Phe Lys Val Arg Met Glu Gly																20
5 10 15 20																
acg gtc aat ggg cac gag ttt gaa ata aaa ggc gaa ggt gaa ggg agg																152
Thr Val Asn Gly His Glu Phe Glu Ile Lys Gly Glu Gly Glu Gly Arg																35
25 30 35																
cct tac gaa ggt cac tgt tcc gta aag ctt atg gta acc aag ggt gga																200
Pro Tyr Glu Gly His Cys Ser Val Lys Leu Met Val Thr Lys Gly Gly																

-24-

40										45										50										
cct	ttg	cca	ttt	gct	ttt	gat	att	ttg	tca	cca	caa	ttt	cag	tat	gga	248														
Pro	Leu	Pro	Phe	Ala	Phe	Asp	Ile	Leu	Ser	Pro	Gln	Phe	Gln	Tyr	Gly															
		55					60					65																		
agc	aag	gta	tat	gtc	aaa	cac	cct	gcc	gac	ata	cca	gac	tat	aaa	aag	296														
Ser	Lys	Val	Tyr	Val	Lys	His	Pro	Ala	Asp	Ile	Pro	Asp	Tyr	Lys	Lys															
	70					75					80																			
ctg	tca	ttt	cct	gag	gga	ttt	aaa	tgg	gaa	agg	gtc	atg	aac	ttt	gaa	344														
Leu	Ser	Phe	Pro	Glu	Gly	Phe	Lys	Trp	Glu	Arg	Val	Met	Asn	Phe	Glu															
	85				90				95						100															
gac	ggt	ggc	gtg	gtt	act	gta	tcc	caa	gat	tcc	agt	ttg	aaa	gac	ggc	392														
Asp	Gly	Gly	Val	Val	Thr	Val	Ser	Gln	Asp	Ser	Ser	Leu	Lys	Asp	Gly															
				105				110						115																
tgt	ttc	atc	tac	gag	gtc	aag	ttc	att	ggg	gtg	aac	ttt	cct	tct	gat	440														
Cys	Phe	Ile	Tyr	Glu	Val	Lys	Phe	Ile	Gly	Val	Asn	Phe	Pro	Ser	Asp															
			120					125					130																	
gga	cct	gtt	atg	cag	agg	agg	aca	cgg	ggc	tgg	gaa	gcc	agc	tct	gag	488														
Gly	Pro	Val	Met	Gln	Arg	Arg	Thr	Arg	Gly	Trp	Glu	Ala	Ser	Ser	Glu															
		135					140					145																		
cgt	ttg	tat	cct	cgt	gat	ggg	gtg	ctg	aaa	gga	gac	atc	cat	atg	gct	536														
Arg	Leu	Tyr	Pro	Arg	Asp	Gly	Val	Leu	Lys	Gly	Asp	Ile	His	Met	Ala															
	150				155						160																			
ctg	agg	ctg	gaa	gga	ggc	ggc	cat	tac	ctc	gtt	gaa	ttc	aaa	agt	att	584														
Leu	Arg	Leu	Glu	Gly	Gly	Gly	His	Tyr	Leu	Val	Glu	Phe	Lys	Ser	Ile															
	165			170						175					180															
tac	atg	gta	aag	aag	cct	tca	gtg	cag	ttg	cca	ggc	tac	tat	tat	gtt	632														
Tyr	Met	Val	Lys	Lys	Pro	Ser	Val	Gln	Leu	Pro	Gly	Tyr	Tyr	Tyr	Val															
				185				190						195																
gac	tcc	aaa	ctg	gat	atg	acg	agc	cac	aac	gaa	gat	tac	aca	gtc	gtt	680														
Asp	Ser	Lys	Leu	Asp	Met	Thr	Ser	His	Asn	Glu	Asp	Tyr	Thr	Val	Val															
		200						205					210																	
gag	cag	tat	gaa	aaa	acc	cag	gga	cgc	cac	cat	ccg	ttc	att	aag	cct	728														
Glu	Gln	Tyr	Glu	Lys	Thr	Gln	Gly	Arg	His	His	Pro	Phe	Ile	Lys	Pro															
		215				220					225																			
ctg	cag	tga	actcggtca	gtcatggatt	agcggtaatg	gccacaaaag										777														
Leu	Gln	*																												
	230																													
gcacgatgat	cgtttttttag	gaatgcagcc	aaaaattgaa	ggttatgaca	gtagaaatac	837																								
aagcaacagg	ctttgcttat	taaacatgta	attgaaaac			876																								

<210> 40

<211> 230

<212> PRT

<213> Discosoma species

<400> 40

Met	Ser	Cys	Ser	Lys	Asn	Val	Ile	Lys	Glu	Phe	Met	Arg	Phe	Lys	Val
1				5					10					15	
Arg	Met	Glu	Gly	Thr	Val	Asn	Gly	His	Glu	Phe	Glu	Ile	Lys	Gly	Glu
			20					25					30		
Gly	Glu	Gly	Arg	Pro	Tyr	Glu	Gly	His	Cys	Ser	Val	Lys	Leu	Met	Val
		35					40					45			
Thr	Lys	Gly	Gly	Pro	Leu	Pro	Phe	Ala	Phe	Asp	Ile	Leu	Ser	Pro	Gln
	50					55				60					

-25-

```

Phe Gln Tyr Gly Ser Lys Val Tyr Val Lys His Pro Ala Asp Ile Pro
65      70      75      80
Asp Tyr Lys Lys Leu Ser Phe Pro Glu Gly Phe Lys Trp Glu Arg Val
      85      90      95
Met Asn Phe Glu Asp Gly Gly Val Val Thr Val Ser Gln Asp Ser Ser
      100      105      110
Leu Lys Asp Gly Cys Phe Ile Tyr Glu Val Lys Phe Ile Gly Val Asn
      115      120      125
Phe Pro Ser Asp Gly Pro Val Met Gln Arg Arg Thr Arg Gly Trp Glu
      130      135      140
Ala Ser Ser Glu Arg Leu Tyr Pro Arg Asp Gly Val Leu Lys Gly Asp
145      150      155      160
Ile His Met Ala Leu Arg Leu Glu Gly Gly His Tyr Leu Val Glu
      165      170      175
Phe Lys Ser Ile Tyr Met Val Lys Lys Pro Ser Val Gln Leu Pro Gly
      180      185      190
Tyr Tyr Tyr Val Asp Ser Lys Leu Asp Met Thr Ser His Asn Glu Asp
      195      200      205
Tyr Thr Val Val Glu Gln Tyr Glu Lys Thr Gln Gly Arg His His Pro
      210      215      220
Phe Ile Lys Pro Leu Gln
225      230

```

```

<210> 41
<211> 25
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> m-att;

```

```

<221> misc_difference
<222> 18
<223> n is a or g or c or t/u

```

```

<400> 41
rkycwgcttt yktrtacnaa stsgb

```

25

```

<210> 42
<211> 25
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> m-attB;

```

```

<221> misc_difference
<222> 18
<223> n is a or g or c or t/u

```

```

<400> 42
agccwgcttt yktrtacnaa ctsgb

```

25

```

<210> 43
<211> 25
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> m-attR

```

```

<221> misc_difference
<222> 18
<223> n is a or g or c or t/u

```

-26-

<400> 43
gttcagcttt cktrtacnaa ctsgb 25

<210> 44
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> m-attL

<221> misc_difference ,
<222> 18
<223> n is a or g or c or t/u

<400> 44
agccwgcttt cktrtacnaa gtsgb 25

<210> 45
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> m-attP1

<221> misc_difference
<222> 18
<223> n is a or g or c or t/u

<400> 45
gttcagcttt yktrtacnaa gtsgb 25

<210> 46
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attB1

<400> 46
agcctgcttt tttgtacaaa cttgt 25

<210> 47
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attB2

<400> 47
agcctgcttt cttgtacaaa cttgt 25

<210> 48
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attB3

<400> 48
accagcttt cttgtacaaa cttgt 25

<210> 49

-27-

<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attR1

<400> 49
gttcagcttt tttgtacaaa cttgt 25

<210> 50
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attR2

<400> 50
gttcagcttt cttgtacaaa cttgt 25

<210> 51
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attR3

<400> 51
gttcagcttt cttgtacaaa gttgg 25

<210> 52
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attL1

<400> 52
agcctgcttt tttgtacaaa gttgg 25

<210> 53
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attL2

<400> 53
agcctgcttt cttgtacaaa gttgg 25

<210> 54
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attL3

<400> 54
accagcttt cttgtacaaa gttgg 25

<210> 55
<211> 25

-28-

<212> DNA
 <213> Artificial Sequence
 <220>
 <223> attP1
 <400> 55
 gttcagcttt tttgtacaaa gttgg 25
 <210> 56
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> attP2,P3
 <400> 56
 gttcagcttt cttgtacaaa gttgg 25
 <210> 57
 <211> 34
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Lox P site
 <400> 57
 ataacttcgt ataatgtatg ctatacgaag ttat 34
 <210> 58
 <211> 1032
 <212> DNA
 <213> Escherichia coli
 <220>
 <221> CDS
 <222> (1)...(1032)
 <223> nucleotide sequence encoding Cre recombinase
 <400> 58
 atg tcc aat tta ctg acc gta cac caa aat ttg cct gca tta ccg gtc 48
 Met Ser Asn Leu Leu Thr Val His Gln Asn Leu Pro Ala Leu Pro Val
 1 5 10 15
 gat gca acg agt gat gag gtt cgc aag aac ctg atg gac atg ttc agg 96
 Asp Ala Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Met Phe Arg
 20 25 30
 gat cgc cag gcg ttt tct gag cat acc tgg aaa atg ctt ctg tcc gtt 144
 Asp Arg Gln Ala Phe Ser Glu His Thr Trp Lys Met Leu Leu Ser Val
 35 40 45
 tgc cgg tcg tgg gcg gca tgg tgc aag ttg aat aac cgg aaa tgg ttt 192
 Cys Arg Ser Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Trp Phe
 50 55 60
 ccc gca gaa cct gaa gat gtt cgc gat tat ctt cta tat ctt cag gcg 240
 Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu Tyr Leu Gln Ala
 65 70 75 80
 cgc ggt ctg gca gta aaa act atc cag caa cat ttg ggc cag cta aac 288
 Arg Gly Leu Ala Val Lys Thr Ile Gln Gln His Leu Gly Gln Leu Asn
 85 90 95
 atg ctt cat cgt cgg tcc ggg ctg cca cga cca agt gac agc aat gct 336

-29-

Met	Leu	His	Arg	Arg	Ser	Gly	Leu	Pro	Arg	Pro	Ser	Asp	Ser	Asn	Ala		
			100					105					110				
ggt	tca	ctg	ggt	atg	cgg	cgg	atc	cga	aaa	gaa	aac	ggt	gat	gcc	ggt	384	
Val	Ser	Leu	Val	Met	Arg	Arg	Ile	Arg	Lys	Glu	Asn	Val	Asp	Ala	Gly		
		115					120					125					
gaa	cgt	gca	aaa	cag	gct	cta	gcg	ttc	gaa	cgc	act	gat	ttc	gac	cag	432	
Glu	Arg	Ala	Lys	Gln	Ala	Leu	Ala	Phe	Glu	Arg	Thr	Asp	Phe	Asp	Gln		
	130					135					140						
ggt	cgt	tca	ctc	atg	gaa	aat	agc	gat	cgc	tgc	cag	gat	ata	cgt	aat	480	
Val	Arg	Ser	Leu	Met	Glu	Asn	Ser	Asp	Arg	Cys	Gln	Asp	Ile	Arg	Asn		
	145				150					155					160		
ctg	gca	ttt	ctg	ggg	att	gct	tat	aac	acc	ctg	tta	cgt	ata	gcc	gaa	528	
Leu	Ala	Phe	Leu	Gly	Ile	Ala	Tyr	Asn	Thr	Leu	Leu	Arg	Ile	Ala	Glu		
				165					170					175			
att	gcc	agg	atc	agg	ggt	aaa	gat	atc	tca	cgt	act	gac	ggt	ggg	aga	576	
Ile	Ala	Arg	Ile	Arg	Val	Lys	Asp	Ile	Ser	Arg	Thr	Asp	Gly	Gly	Arg		
			180					185					190				
atg	tta	atc	cat	att	ggc	aga	acg	aaa	acg	ctg	ggt	agc	acc	gca	ggt	624	
Met	Leu	Ile	His	Ile	Gly	Arg	Thr	Lys	Thr	Leu	Val	Ser	Thr	Ala	Gly		
		195					200					205					
gta	gag	aag	gca	ctt	agc	ctg	ggg	gta	act	aaa	ctg	gtc	gag	cga	tgg	672	
Val	Glu	Lys	Ala	Leu	Ser	Leu	Gly	Val	Thr	Lys	Leu	Val	Glu	Arg	Trp		
	210					215					220						
att	tcc	gtc	tct	ggt	gta	gct	gat	gat	ccg	aat	aac	tac	ctg	ttt	tgc	720	
Ile	Ser	Val	Ser	Gly	Val	Ala	Asp	Asp	Pro	Asn	Asn	Tyr	Leu	Phe	Cys		
	225				230					235					240		
cgg	gtc	aga	aaa	aat	ggt	ggt	gcc	gcg	cca	tct	gcc	acc	agc	cag	cta	768	
Arg	Val	Arg	Lys	Asn	Gly	Val	Ala	Ala	Pro	Ser	Ala	Thr	Ser	Gln	Leu		
				245					250					255			
tca	act	cgc	gcc	ctg	gaa	ggg	att	ttt	gaa	gca	act	cat	cga	ttg	att	816	
Ser	Thr	Arg	Ala	Leu	Glu	Gly	Ile	Phe	Glu	Ala	Thr	His	Arg	Leu	Ile		
			260					265					270				
tac	ggc	gct	aag	gat	gac	tct	ggt	cag	aga	tac	ctg	gcc	tgg	tct	gga	864	
Tyr	Gly	Ala	Lys	Asp	Asp	Ser	Gly	Gln	Arg	Tyr	Leu	Ala	Trp	Ser	Gly		
		275					280					285					
cac	agt	gcc	cgt	gtc	gga	gcc	gcg	cga	gat	atg	gcc	cgc	gct	gga	ggt	912	
His	Ser	Ala	Arg	Val	Gly	Ala	Ala	Arg	Asp	Met	Ala	Arg	Ala	Gly	Val		
	290					295					300						
tca	ata	ccg	gag	atc	atg	caa	gct	ggt	ggc	tgg	acc	aat	gta	aat	att	960	
Ser	Ile	Pro	Glu	Ile	Met	Gln	Ala	Gly	Gly	Trp	Thr	Asn	Val	Asn	Ile		
	305				310					315					320		
gtc	atg	aac	tat	atc	cgt	aac	ctg	gat	agt	gaa	aca	ggg	gca	atg	gtg	1008	
Val	Met	Asn	Tyr	Ile	Arg	Asn	Leu	Asp	Ser	Glu	Thr	Gly	Ala	Met	Val		
				325					330					335			
cgc	ctg	ctg	gaa	gat	ggc	gat	tag									1032	
Arg	Leu	Leu	Glu	Asp	Gly	Asp	*										
			340														

<210> 59
 <211> 343
 <212> PRT

-30-

<213> *Escherichia coli*

<400> 59

```

Met Ser Asn Leu Leu Thr Val His Gln Asn Leu Pro Ala Leu Pro Val
1      5      10      15
Asp Ala Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Met Phe Arg
      20      25      30
Asp Arg Gln Ala Phe Ser Glu His Thr Trp Lys Met Leu Leu Ser Val
      35      40      45
Cys Arg Ser Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Trp Phe
      50      55      60
Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu Tyr Leu Gln Ala
      65      70      75      80
Arg Gly Leu Ala Val Lys Thr Ile Gln Gln His Leu Gly Gln Leu Asn
      85      90      95
Met Leu His Arg Arg Ser Gly Leu Pro Arg Pro Ser Asp Ser Asn Ala
      100      105      110
Val Ser Leu Val Met Arg Arg Ile Arg Lys Glu Asn Val Asp Ala Gly
      115      120      125
Glu Arg Ala Lys Gln Ala Leu Ala Phe Glu Arg Thr Asp Phe Asp Gln
      130      135      140
Val Arg Ser Leu Met Glu Asn Ser Asp Arg Cys Gln Asp Ile Arg Asn
      145      150      155      160
Leu Ala Phe Leu Gly Ile Ala Tyr Asn Thr Leu Leu Arg Ile Ala Glu
      165      170      175
Ile Ala Arg Ile Arg Val Lys Asp Ile Ser Arg Thr Asp Gly Gly Arg
      180      185      190
Met Leu Ile His Ile Gly Arg Thr Lys Thr Leu Val Ser Thr Ala Gly
      195      200      205
Val Glu Lys Ala Leu Ser Leu Gly Val Thr Lys Leu Val Glu Arg Trp
      210      215      220
Ile Ser Val Ser Gly Val Ala Asp Asp Pro Asn Asn Tyr Leu Phe Cys
      225      230      235      240
Arg Val Arg Lys Asn Gly Val Ala Ala Pro Ser Ala Thr Ser Gln Leu
      245      250      255
Ser Thr Arg Ala Leu Glu Gly Ile Phe Glu Ala Thr His Arg Leu Ile
      260      265      270
Tyr Gly Ala Lys Asp Asp Ser Gly Gln Arg Tyr Leu Ala Trp Ser Gly
      275      280      285
His Ser Ala Arg Val Gly Ala Ala Arg Asp Met Ala Arg Ala Gly Val
      290      295      300
Ser Ile Pro Glu Ile Met Gln Ala Gly Gly Trp Thr Asn Val Asn Ile
      305      310      315      320
Val Met Asn Tyr Ile Arg Asn Leu Asp Ser Glu Thr Gly Ala Met Val
      325      330      335
Arg Leu Leu Glu Asp Gly Asp
      340

```

<210> 60

<211> 1272

<212> DNA

<213> *Saccharomyces cerevisiae*

<220>

<221> CDS

<222> (1)...(1272)

<223> nucleotide sequence encoding Flip recombinase

<400> 60

```

atg cca caa ttt ggt ata tta tgt aaa aca cca cct aag gtg ctt gtt      48
Met Pro Gln Phe Gly Ile Leu Cys Lys Thr Pro Pro Lys Val Leu Val
1      5      10
cgt cag ttt gtg gaa agg ttt gaa aga cct tca ggt gag aaa ata gca      96
Arg Gln Phe Val Glu Arg Phe Glu Arg Pro Ser Gly Glu Lys Ile Ala
      20      25      30

```

-31-

tta	tgt	gct	gct	gaa	cta	acc	tat	tta	tgt	tgg	atg	att	aca	cat	aac	144
Leu	Cys	Ala	Ala	Glu	Leu	Thr	Tyr	Leu	Cys	Trp	Met	Ile	Thr	His	Asn	
		35					40					45				
gga	aca	gca	atc	aag	aga	gcc	aca	ttc	atg	agc	tat	aat	act	atc	ata	192
Gly	Thr	Ala	Ile	Lys	Arg	Ala	Thr	Phe	Met	Ser	Tyr	Asn	Thr	Ile	Ile	
	50					55					60					
agc	aat	tcg	ctg	agt	ttc	gat	att	gtc	aat	aaa	tca	ctc	cag	ttt	aaa	240
Ser	Asn	Ser	Leu	Ser	Phe	Asp	Ile	Val	Asn	Lys	Ser	Leu	Gln	Phe	Lys	
	65				70					75					80	
tac	aag	acg	caa	aaa	gca	aca	att	ctg	gaa	gcc	tca	tta	aag	aaa	ttg	288
Tyr	Lys	Thr	Gln	Lys	Ala	Thr	Ile	Leu	Glu	Ala	Ser	Leu	Lys	Lys	Leu	
				85					90					95		
att	cct	gct	tgg	gaa	ttt	aca	att	att	cct	tac	tat	gga	caa	aaa	cat	336
Ile	Pro	Ala	Trp	Glu	Phe	Thr	Ile	Ile	Pro	Tyr	Tyr	Gly	Gln	Lys	His	
			100					105					110			
caa	tct	gat	atc	act	gat	att	gta	agt	agt	ttg	caa	tta	cag	ttc	gaa	384
Gln	Ser	Asp	Ile	Thr	Asp	Ile	Val	Ser	Ser	Leu	Gln	Leu	Gln	Phe	Glu	
		115					120					125				
tca	tcg	gaa	gaa	gca	gat	aag	gga	aat	agc	cac	agt	aaa	aaa	atg	ctt	432
Ser	Ser	Glu	Glu	Ala	Asp	Lys	Gly	Asn	Ser	His	Ser	Lys	Lys	Met	Leu	
	130					135					140					
aaa	gca	ctt	cta	agt	gag	ggg	gaa	agc	atc	tgg	gag	atc	act	gag	aaa	480
Lys	Ala	Leu	Leu	Ser	Glu	Gly	Glu	Ser	Ile	Trp	Glu	Ile	Thr	Glu	Lys	
	145				150					155					160	
ata	cta	aat	tcg	ttt	gag	tat	act	tcg	aga	ttt	aca	aaa	aca	aaa	act	528
Ile	Leu	Asn	Ser	Phe	Glu	Tyr	Thr	Ser	Arg	Phe	Thr	Lys	Thr	Lys	Thr	
				165					170					175		
tta	tac	caa	ttc	ctc	ttc	cta	gct	act	ttc	atc	aat	tgt	gga	aga	ttc	576
Leu	Tyr	Gln	Phe	Leu	Phe	Leu	Ala	Thr	Phe	Ile	Asn	Cys	Gly	Arg	Phe	
			180					185					190			
agc	gat	att	aag	aac	gtt	gat	ccg	aaa	tca	ttt	aaa	tta	gtc	caa	aat	624
Ser	Asp	Ile	Lys	Asn	Val	Asp	Pro	Lys	Ser	Phe	Lys	Val	Val	Gln	Asn	
		195					200					205				
aag	tat	ctg	gga	gta	ata	atc	cag	tgt	tta	gtg	aca	gag	aca	aag	aca	672
Lys	Tyr	Leu	Gly	Val	Ile	Ile	Gln	Cys	Leu	Val	Thr	Glu	Thr	Lys	Thr	
	210					215					220					
agc	gtt	agt	agg	cac	ata	tac	ttc	ttt	agc	gca	agg	ggg	agg	atc	gat	720
Ser	Val	Ser	Arg	His	Ile	Tyr	Phe	Phe	Ser	Ala	Arg	Gly	Arg	Ile	Asp	
	225				230					235					240	
cca	ctt	gta	tat	ttg	gat	gaa	ttt	ttg	agg	aat	tct	gaa	cca	gtc	cta	768
Pro	Leu	Val	Tyr	Leu	Asp	Glu	Phe	Leu	Arg	Asn	Ser	Glu	Pro	Val	Leu	
				245					250					255		
aaa	cga	gta	aat	agg	acc	ggc	aat	tct	tca	agc	aat	aaa	cag	gaa	tac	816
Lys	Arg	Val	Asn	Arg	Thr	Gly	Asn	Ser	Ser	Ser	Asn	Lys	Gln	Glu	Tyr	
			260					265					270			
caa	tta	tta	aaa	gat	aac	tta	gtc	aga	tcg	tac	aat	aaa	gct	ttg	aag	864
Gln	Leu	Leu	Lys	Asp	Asn	Leu	Val	Arg	Ser	Tyr	Asn	Lys	Ala	Leu	Lys	
		275					280					285				
aaa	aat	gcg	cct	tat	tca	atc	ttt	gct	ata	aaa	aat	ggc	cca	aaa	tct	912
Lys	Asn	Ala	Pro	Tyr	Ser	Ile	Phe	Ala	Ile	Lys	Asn	Gly	Pro	Lys	Ser	
	290					295					300					

-32-

cac His 305	att Ile	gga Gly	aga Arg	cat His	ttg Leu 310	atg Met	acc Thr	tca Ser	ttt Phe	ctt Leu 315	tca Ser	atg Met	aag Lys	ggc Gly	cta Leu 320	960
acg Thr	gag Glu	ttg Leu	act Thr	aat Asn 325	gtt Val	gtg Val	gga Gly	aat Asn 330	tgg Trp	agc Ser	gat Asp	aag Lys	cgt Arg	gct Ala 335	tct Ser	1008
gcc Ala	gtg Val	gcc Ala	agg Arg 340	aca Thr	acg Thr	tat Tyr	act Thr	cat His 345	cag Gln	ata Ile	aca Thr	gca Ala	ata Ile 350	cct Pro	gat Asp	1056
cac His	tac Tyr	ttc Phe 355	gca Ala	cta Leu	gtt Val	tct Ser	cgg Arg 360	tac Tyr	tat Tyr	gca Ala	tat Tyr	gat Asp 365	cca Pro	ata Ile	tca Ser	1104
aag Lys 370	gaa Glu	atg Met	ata Ile	gca Ala	ttg Leu	aag Lys 375	gat Asp	gag Glu	act Thr	aat Asn 380	cca Pro	att Ile	gag Glu	gag Glu	tgg Trp	1152
cag Gln 385	cat His	ata Ile	gaa Glu	cag Gln	cta Leu 390	aag Lys	ggg Gly	agt Ser	gct Ala	gaa Glu 395	gga Gly	agc Ser	ata Ile	cga Arg	tac Tyr 400	1200
ccc Pro	gca Ala	tgg Trp	aat Asn	ggg Gly 405	ata Ile	ata Ile	tca Ser	cag Gln	gag Glu 410	gta Val	cta Leu	gac Asp	tac Tyr	ctt Leu 415	tca Ser	1248
tcc Ser	tac Tyr	ata Ile	aat Asn	aga Arg	cgc Arg	ata Ile	taa *									1272

<210> 61

<211> 422

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 61

Pro 1	Gln	Phe	Gly	Ile 5	Leu	Cys	Lys	Thr	Pro 10	Pro	Lys	Val	Leu 15	Val	Arg
Gln	Phe	Val	Glu	Arg	Phe	Glu	Arg	Pro 25	Ser	Gly	Glu	Lys	Ile 30	Ala	Leu
Cys	Ala	Ala	Glu	Leu	Thr	Tyr	Leu 40	Cys	Trp	Met	Ile	Thr 45	His	Asn	Gly
Thr	Ala	Ile	Lys	Arg	Ala	Thr	Phe 55	Met	Ser	Tyr	Asn 60	Thr	Ile	Ile	Ser
Asn 65	Ser	Leu	Ser	Phe	Asp 70	Ile	Val	Asn	Lys	Ser 75	Leu	Gln	Phe	Lys	Tyr 80
Lys	Thr	Gln	Lys	Ala	Thr 85	Ile	Leu	Glu	Ala 90	Ser	Leu	Lys	Lys	Leu	Ile 95
Pro	Ala	Trp	Glu	Phe	Thr	Ile	Ile	Pro 105	Tyr	Tyr	Gly	Gln	Lys 110	His	Gln
Ser	Asp	Ile	Thr	Asp	Ile	Val	Ser 120	Ser	Leu	Gln	Leu	Gln 125	Phe	Glu	Ser
Ser	Glu	Glu	Ala	Asp	Lys	Gly	Asn 135	Ser	His	Ser	Lys	Lys	Met	Leu	Lys
Ala 145	Leu	Leu	Ser	Glu	Gly 150	Glu	Ser	Ile	Trp	Glu 155	Ile	Thr	Glu	Lys	Ile 160
Leu	Asn	Ser	Phe	Glu 165	Tyr	Thr	Ser	Arg	Phe 170	Thr	Lys	Thr	Lys	Thr 175	Leu
Tyr	Gln	Phe	Leu	Phe 180	Leu	Ala	Thr	Phe 185	Ile	Asn	Cys	Gly	Arg	Phe	Ser 190
Asp	Ile	Lys	Asn	Val	Asp	Pro	Lys 200	Ser	Phe	Lys	Leu	Val	Gln	Asn	Lys
Tyr	Leu	Gly	Val	Ile	Ile	Gln	Cys	Leu	Val	Thr	Glu	Thr	Lys	Thr	Ser

-33-

210	Val	Ser	Arg	His	Ile	Tyr	215	Phe	Phe	Ser	Ala	Arg	220	Gly	Arg	Ile	Asp	Pro
225	Leu	Val	Tyr	Leu	Asp	Glu	230	Phe	Leu	Arg	Asn	Ser	235	Glu	Pro	Val	Leu	Lys
				245	Thr	Gly	Asn	Ser	Ser	Ser	Asn	Lys	250	Gln	Glu	Tyr	Gln	
	Arg	Val	Asn	Arg	260	Leu	Val	Arg	Ser	Tyr	Asn	Lys	265	Ala	Leu	Lys	Lys	
	Leu	Leu	Lys	Asp	Asn	Leu	Val	Arg	Ser	Tyr	Asn	Lys	270	Ala	Leu	Lys	Lys	
	Asn	Ala	Pro	Tyr	Ser	Ile	Phe	Ala	Ile	Lys	Asn	Gly	285	Pro	Lys	Ser	His	
	Ile	Gly	Arg	His	Leu	Met	Thr	Ser	Phe	Leu	Ser	Met	300	Lys	Gly	Leu	Thr	
	305	Glu	Leu	Thr	Asn	Val	Gly	Asn	Trp	Ser	Asp	Lys	315	Arg	Ala	Ser	Ala	
				325	Thr	Tyr	Thr	His	Gln	Ile	Thr	Ala	330	Ile	Pro	Asp	His	
	Val	Ala	Arg	340	Val	Ser	Arg	Tyr	Tyr	Ala	Tyr	Asp	345	Pro	Ile	Ser	Lys	
	Tyr	Phe	Ala	Leu	Val	Ser	Arg	Tyr	Tyr	Ala	Tyr	Asp	355	Pro	Ile	Ser	Lys	
	Glu	Met	Ile	Ala	Leu	Lys	Asp	Glu	Thr	Asn	Pro	Ile	360	Glu	Glu	Trp	Gln	
	370	His	Ile	Glu	Gln	Leu	Lys	Gly	Ser	Ala	Glu	Gly	375	Ser	Ile	Arg	Tyr	
	385	Ala	Trp	Asn	Gly	Ile	Ile	Ser	Gln	Glu	Val	Leu	380	Asp	Tyr	Leu	Ser	
				405	Arg	Ile					410		395					
	Tyr	Ile	Asn	Arg	420								415					

<210> 62
 <211> 48
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> IR2

<400> 62
 gaagttccta ttccgaagtt cctattctct agaaagtata ggaacttc

48

<210> 63
 <211> 48
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> IR1

<400> 63
 gaagttccta tactttctag agaataggaa cttcggaata ggaacttc

48

<210> 64
 <211> 66
 <212> DNA
 <213> Bacteriophage mu

<220>
 <221> CDS
 <222> (1)...(66)
 <223> nucleotide sequence encoding GIN recombinase

<400> 64
 tca act ctg tat aaa aaa cac ccc gcg aaa cga gcg cat ata gaa aac
 Ser Thr Leu Tyr Lys Lys His Pro Ala Lys Arg Ala His Ile Glu Asn
 1 5 10 15

48

gac gat cga atc aat taa
 Asp Asp Arg Ile Asn *

66

-34-

20

<210> 65
 <211> 21
 <212> PRT
 <213> bacteriophage mu

<400> 65
 Ser Thr Leu Tyr Lys Lys His Pro Ala Lys Arg Ala His Ile Glu Asn
 1 5 10 15
 Asp Asp Arg Ile Asn
 20

<210> 66
 <211> 69
 <212> DNA
 <213> Bacteriophage mu

<220>
 <221> CDS
 <222> (1)...(69)
 <223> nucleotide sequence encoding Gin recombinase

<400> 66
 tat aaa aaa cat ccc gcg aaa cga acg cat ata gaa aac gac gat cga 48
 Tyr Lys Lys His Pro Ala Lys Arg Thr His Ile Glu Asn Asp Asp Arg
 1 5 10 15
 atc aat caa atc gat cgg taa 69
 Ile Asn Gln Ile Asp Arg *
 20

<210> 67
 <211> 22
 <212> PRT
 <213> bacteriophage mu

<220>
 <223> Gin recombinase of bacteriophage mu

<400> 67
 Tyr Lys Lys His Pro Ala Lys Arg Thr His Ile Glu Asn Asp Asp Arg
 1 5 10 15
 Ile Asn Gln Ile Asp Arg
 20

<210> 68
 <211> 555
 <212> DNA
 <213> Escherichia coli

<220>
 <221> CDS
 <222> (1)...(555)
 <223> nucleotide sequence encoding PIN recombinase

<400> 68
 atg ctt att ggc tat gta cgc gta tca aca aat gac cag aac aca gat 48
 Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
 1 5 10 15
 cta caa cgt aat gcg ctg aac tgt gca gga tgc gag ctg att ttt gaa 96
 Leu Gln Arg Asn Ala Leu Asn Cys Ala Gly Cys Glu Leu Ile Phe Glu
 20 25 30

-35-

gac aag ata agc ggc aca aag tcc gaa agg ccg gga ctg aaa aaa ctg	144
Asp Lys Ile Ser Gly Thr Lys Ser Glu Arg Pro Gly Leu Lys Lys Leu	
35 40 45	
ctc agg aca tta tcc gca ggt gac act ctg gtt gtc tgg aag ctg gat	192
Leu Arg Thr Leu Ser Ala Gly Asp Thr Leu Val Val Trp Lys Leu Asp	
50 55 60	
cgg ctg ggg cgt agt atg cgg cat ctt gtc gtg ctg gtg gag gag ttg	240
Arg Leu Gly Arg Ser Met Arg His Leu Val Val Leu Val Glu Glu Leu	
65 70 75 80	
cgc gaa cga ggc atc aac ttt cgt agt ctg acg gat tca att gat acc	288
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr	
85 90 95	
agc aca cca atg gga cgc ttt ttc ttt cat gtg atg ggt gcc ctg gct	336
Ser Thr Pro Met Gly Arg Phe Phe Phe His Val Met Gly Ala Leu Ala	
100 105 110	
gaa atg gag cgt gaa ctg att gtt gaa cga aca aaa gct gga ctg gaa	384
Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Lys Ala Gly Leu Glu	
115 120 125	
act gct cgt gca cag gga cga att ggt gga cgt cgt ccc aaa ctt aca	432
Thr Ala Arg Ala Gln Gly Arg Ile Gly Gly Arg Arg Pro Lys Leu Thr	
130 135 140	
cca gaa caa tgg gca caa gct gga cga tta att gca gca gga act cct	480
Pro Glu Gln Trp Ala Gln Ala Gly Arg Leu Ile Ala Ala Gly Thr Pro	
145 150 155 160	
cgc cag aag gtg gcg att atc tat gat gtt ggt gtg tca act ttg tat	528
Arg Gln Lys Val Ala Ile Ile Tyr Asp Val Gly Val Ser Thr Leu Tyr	
165 170 175	
aag agg ttt cct gca ggg gat aaa taa	555
Lys Arg Phe Pro Ala Gly Asp Lys *	
180	

<210> 69

<211> 184

<212> PRT

<213> Escherichia coli

<400> 69

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp	
1 5 10 15	
Leu Gln Arg Asn Ala Leu Asn Cys Ala Gly Cys Glu Leu Ile Phe Glu	
20 25 30	
Asp Lys Ile Ser Gly Thr Lys Ser Glu Arg Pro Gly Leu Lys Lys Leu	
35 40 45	
Leu Arg Thr Leu Ser Ala Gly Asp Thr Leu Val Val Trp Lys Leu Asp	
50 55 60	
Arg Leu Gly Arg Ser Met Arg His Leu Val Val Leu Val Glu Glu Leu	
65 70 75 80	
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr	
85 90 95	
Ser Thr Pro Met Gly Arg Phe Phe Phe His Val Met Gly Ala Leu Ala	
100 105 110	
Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Lys Ala Gly Leu Glu	
115 120 125	
Thr Ala Arg Ala Gln Gly Arg Ile Gly Gly Arg Arg Pro Lys Leu Thr	
130 135 140	
Pro Glu Gln Trp Ala Gln Ala Gly Arg Leu Ile Ala Ala Gly Thr Pro	
145 150 155 160	

-36-

Arg Gln Lys Val Ala Ile Ile Tyr Asp Val Gly Val Ser Thr Leu Tyr
 165 170 175
 Lys Arg Phe Pro Ala Gly Asp Lys
 180

<210> 70
 <211> 4778
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pcx plasmid

<400> 70
 gtcgacattg attatttgact agttattaat agtaatcaat tacgggggtca ttagttcata 60
 gcccataatat ggagttccgc gttacataac ttacggtaaa tggcccgccct ggctgaccgc 120
 ccaacgaccc ccgcccatg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180
 ggacttttcca ttgagctcaa tgggtggact atttacggta aactgcccac ttggcagtac 240
 atcaagtgtat tcatatgccca agtacgcccc ctattgacgt caatgacggt aaatggcccg 300
 cctggcatta tgcccagtac atgaccttat gggactttcc tacttggcag tacatctacg 360
 tattagtcac cgctattacc atgggtcgag gtgagcccca cgttctgctt cactctcccc 420
 atctcccccc cctccccacc cccaattttg tatttattta ttttttaatt attttgtgca 480
 gcgatggggg cggggggggg gggggcgcg cccaggcggg gcggggcggg gcgagggggc 540
 gggcgggggc aggcggagag gtgcggcggc agccaatcag agcggcgcg cccgaaagt 600
 tcctttttatg gcgagggcg ggcggcgggc gccctataaa aagcgaagcg cgcggcgggc 660
 gggagtcgct gcgttgccct cggcccgctgc cccgctccgc gccgcctcgc gccgcccgc 720
 ccgctctga ctgaccgct tacccccaca ggtgagcggg cgggacggcc cttctcctcc 780
 gggctgtaat tagcgcttgg tttaatgacg gctcgtttct tttctgtggc tgcgtgaaag 840
 ccttaaaggg ctccgggagg gccctttgtg cggggggggg cggctcgggg ggtgctgctg 900
 tgtgtgtgtg cgtggggagc gcccgctgcg gcccgcctg cccggcggt gtgagcgtg 960
 cgggcgcggc gcggggctt gtgcgctccg cgtgtgcgag aggggagcgc ggccgggggc 1020
 ggtgccccgc ggtgcggggg ggctgcgagg ggaacaaagg ctgctgctgc ggtgtgtgct 1080
 tgggggggtg agcaggggtg gtggcgcgcg cggtcgggct gtaaccccc cctgcacccc 1140
 cctccccgag ttgctgagca cggcccggtc tcgggtgcgg ggctccgtgc ggggctggc 1200
 gcggggctcg ccgtgcccgg cgggggggtg cggcaggtg ggggtgccgg cggggcggg 1260
 ccgctcggg ctggggaggg ctcgggggag gggcgcggtg gccccggagc gccggcggt 1320
 gtcgagggcg gcgagccgc agccattgcc ttttatggta atcgtgcgag agggcgccgc 1380
 gacttccttt gtcccaaacc tggcggagcc gaaatctggg aggcgcggcc gcacccctc 1440
 tagcggcgcg ggcggaagcg gtgcggcgcc ggcaggaagg aaatggcgcg ggagggcctt 1500
 cgtgcgtcgc gcgcgcccg cccctctctc catctccagc ctcggggctg ccgcaggggg 1560
 acggctgcct tcggggggga cggggcaggg cgggggtcgg cttctggcgt gtgaccggcg 1620
 gctctagag catgttcatg ccttctctt tttctacag ttctcactc attcactcct 1680
 acgtgctggt tgttgtgctg tctcatcatt ttggcaaga attcactcct cagggtgcagg 1740
 ctgcctatca gaagggtggt gctggtgtgg ccaatgccct ggctcacaaa taccactgag 1800
 atctttttcc ctctgcaaaa aattatgggg acatcatgaa gcccttgag catctgactt 1860
 ctggctaata aaggaaattt attttcattg caatagtgt ttggaatttt ttgtgtctct 1920
 cactcggaag gacatatggg agggcaaatc atttaaaaca tcagaatgag tatttggttt 1980
 agagtttggc aacatatgcc atatgctggc tgccatgaac aaaggtggct ataaagaggt 2040
 catcagtaata tgaaacagcc cctgctgtgc cactccctat tccatagaaa agccttgact 2100
 tgaggttaga ttttttttat attttgtttt gtgttatttt tttctttaa atccctaaaa 2160
 ttttccttac atgttttact agccagattt ttctctctc cctgactact cccagtcata 2220
 gctgtccctc ttctcttatg aagatccctc gacctgcagc ccaagcttgg cgtaatcatg 2280
 gtcatagctg tttcctgtgt gaaattgtta tccgctcaca attccacaca acatacgagc 2340
 cggaagcata aagtgtaaag cctgggggtg ctaatgagtg agctaactca cattaattgc 2400
 gttgcgtca ctgcccgtt tccagtcggg aaacctgtcg tgccagcgga tccgcatctc 2460
 aattagtcag caaccatagt cccgccccta actccgccc tcccggccct aactccgcc 2520
 agttccggcc atttccggc ccatggctga ctaatttttt ttattttatgc agaggccgag 2580
 gccgctcgg cctctgagct attccagaag tagtgaggag gcttttttgg aggcctaggc 2640
 ttttgcaaaa agctaacttg tttattgcag cttataatgg ttacaaataa agcaatagca 2700
 tcacaaattt cacaaataaa gcattttttt cactgcattc tagttgtggt ttgtccaaac 2760
 tcatcaatgt atcttatcat gtctggatcc gctgcattaa tgaatcgggc aacgcgcggg 2820
 gagaggcggt ttgctgtatt ggcgctcttc cgcttcctcg ctcaactgact cgctgcgctc 2880
 ggtcggtcgg ctgcggcgag cggatcagc tcactcaaaa gcggtaatac ggttatccac 2940
 agaatcaggg gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggcaggaa 3000
 ccgtaaaaag gccgcgttgc tggcggtttt ccataggctc cgccccctg acgagcatca 3060
 caaaaatcga cgctcaagtc agaggtggcg aaaccgcaca ggactataaa gataccaggc 3120
 gtttccccct ggaagctccc tcgtgcgctc tcctgttccg accctgcccg ttaccggata 3180

-37-

cctgtccgcc	ttttccctt	cggaagcgt	ggcgcttct	caatgctcac	gctgtaggta	3240
tctcagttcg	gtgtaggteg	ttcgctccaa	gctgggctgt	gtgcacgaac	cccccgttca	3300
gcccagaccgc	tgcgccttat	ccggttaacta	tgcgtcttgag	tccaacccgg	taagacacga	3360
cttatcgcca	ctggcagcag	ccactggtaa	caggattagc	agagcgaggt	atgtaggcgg	3420
tgctacagag	ttcttgaagt	gggtggcctaa	ctacggctac	actagaagga	cagtatttgg	3480
tatctgcgct	ctgctgaagc	cagttacctt	cggaaaaaga	gttggtagct	cttgatccgg	3540
caaaacaaacc	accgctggta	gcgggtgggtt	ttttgtttgc	aagcagcaga	ttacgcgcag	3600
aaaaaaagga	tctcaagaag	atcctttgat	cttttctacg	gggtctgacg	ctcagtgga	3660
cgaaaactca	cgtaagggga	ttttgggtcat	gagattatca	aaaaggatct	tcacctagat	3720
ccttttaaat	taaaaatgaa	gttttaaatc	aatctaaagt	atatatgagt	aaacttggtc	3780
tgacagttac	caatgcttaa	tcagtgagge	acctatctca	gcgatctgtc	tatttcggtc	3840
atccatagtt	gcctgactcc	ccgtcgtgta	gataactacg	atacgggagg	gcttaccatc	3900
tggtcccgagt	gctgcaatga	taccgcgaga	cccacgctca	ccggtccag	atztatcagc	3960
aataaacagag	ccagccgga	gggcccagcg	cagaagggt	cctgcaactt	tatccgcctc	4020
catccagttc	attaattggt	gccgggaagc	tagagtaagt	agttcgccag	ttaatagttt	4080
gcgcaacggt	gttgccattg	ctacaggcat	cgtgggtgta	cgctcgctgt	ttggtatggc	4140
ttcattcagc	tccggttccc	aacgatcaag	gcgagttaca	tgatcccca	tgatgtgcaa	4200
aaaagcgggt	agctccttcg	gtcctccgat	cgttgtcaga	agtaagttgg	ccgcagtggt	4260
atcactcatg	gttatggcag	cactgcataa	ttctcttact	gtcatgccat	ccgtaagatg	4320
cttttctgtg	actggtgagt	actcaaccac	gtcattctga	gaatagtgt	tcgcccgcac	4380
gagttgctct	tgcccggcgt	caatacggga	taataccgcg	ccacatagca	gaactttaaa	4440
agtgtctcat	attggaaaaac	gttcttcggg	gcgaaaaactc	tcaaggatct	taccgctggt	4500
gagatccagt	tcgatgtaac	ccactcgtgc	acccaactga	tcttcagcat	cttttacttt	4560
caccagcgtt	tctgggtgag	caaaaaacagg	aaggcaaaat	gccgcaaaaa	agggataaag	4620
ggcgacacgg	aaatggtgaa	tactcatact	cttccttttt	caatattatt	gaagcattta	4680
tcaggggttat	tgtctcatga	gcggatacat	atttgaatgt	atttagaaaa	ataaacaatt	4740
aggggttccg	cgcacatttc	cccgaagaat	gccacctg			4778

<210> 71

<211> 5510

<212> DNA

<213> Artificial Sequence

<220>

<223> pCXeGFP plasmid

<400> 71

gtcgacattg	attattgact	agttattaat	agtaatcaat	tacgggggtca	ttagttcata	60
gcccataatat	ggagttccgc	gttacataac	ttacggtaaa	tggtccgcct	ggctgaccgc	120
ccaacgaccc	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	acgccaatag	180
ggactttcca	ttgacgtcaa	tgggtggact	atttacggta	aactgcccac	ttggcagtag	240
atcaagtgt	tcatagtcca	agtacgcccc	ctattgacgt	caatgacggg	aaatggcccc	300
cctggcatta	tgcccagtag	atgaccttat	gggactttcc	tacttggcag	tacatctacg	360
tattagtcat	cgctattacc	atgggtcgag	gtgagcccca	cggtctgtct	cactctcccc	420
atctcccccc	cctccccacc	cccaattttg	tattttatta	tttttttaatt	attttgtgca	480
gcgatggggg	cggggggggg	gggggcgcgc	gccaggcggg	gcggggcggg	gcgagggggc	540
gggcccggcg	aggcggagag	gtgcggcgcc	agccaatcag	agcggcgccg	tccgaaagtt	600
tccttttatg	gcgaggcgcc	ggcggcgggc	gcccataaaa	aagcgaagcg	cgcggcgggc	660
gggagtcgct	gcgttgccct	cgccccgtgc	cccgtccgcg	gccgcctcgc	gccgcccggc	720
ccggctctga	ctgaccgcgt	tactcccaca	ggtagcggg	cgggacggcc	cttctcctcc	780
gggctgtaat	tagcgcttgg	tttaatgacg	gctcgtttct	tttctgtggc	tgctgaaaag	840
ccttaaaggg	ctccgggagg	gccctttgtg	cgggggggag	cggtccgggg	ggtgcgtgct	900
tgtgtgtgtg	cgtggggagc	gccgcgtgag	gcccgcgctg	cccggcggct	gtgagcgctg	960
cgggcgcggc	gcggggcctt	gtgcgctccg	cgtgtgcgcg	aggggagcgc	ggccgggggg	1020
ggtgcccccg	ggtgcggggg	ggctgcgagg	ggaacaaagg	ctgcgtgcgg	ggtgtgtgct	1080
tggggggggt	agcagggggg	gtgggcgcgg	cggtcgggct	gtaaccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cggcccggct	tcgggtgcgg	ggctccgtgc	ggggcgtggc	1200
gcggggcctcg	ccgtgccggg	cgggggggtg	cggcaggtgg	gggtgccggg	cggggcgggg	1260
ccgcctcggg	ccggggaggg	ctcgggggag	gggcgcggcg	gccccggagc	gccggcgggc	1320
gtcgaggcgc	gcgagccgcg	agccattgcc	ttttatggta	atcgtgcgag	agggcgcagg	1380
gacttccttt	gtcccaaatc	tggcggagcc	gaaatctggg	aggcgcggcc	gcaccccctc	1440
tagcggcgcg	gggcgaagcg	gtgcggcgcc	ggcaggaagg	aaatgggcgg	ggagggcctt	1500
cgtgcgtcgc	cgcgcgcgcg	tccccttctc	catctccagc	ctcggggctg	ccgcaggggg	1560
acggctgcct	tccgggggga	cggggcaggg	cgggggttcg	cttctggcgt	gtgaccggcg	1620
gctctagagc	ctctgctaac	catgttcatt	ccttcttctt	tttctctacg	ctcctgggca	1680
acgtgcgtgt	tgttgtgctg	tctcatcatt	ttggcaaaag	attcgccacc	atgggtgagca	1740
agggcgaggga	gctgtttcacc	gggggtggtgc	ccatcctggt	cgagctggac	ggcgacgtaa	1800

-38-

acggccacaa	gttcagcgtg	tccggcgagg	gcgagggcga	tgccacctac	ggcaagctga	1860
ccctgaagtt	catctgcacc	acgggcaagc	tgcccggtgc	ctggccacc	ctcgtagcca	1920
ccctgaccta	cggcggtgcag	tgcttcagcc	gctaccccgga	ccacatgaag	cagcacgact	1980
tcttcaagtc	cgccatgccc	gaagggtacg	tccaggagcg	caccatcttc	ttcaaggacg	2040
acggcaacta	caagaccgcg	gcccaggtga	agttcgaggg	cgacacctg	gtgaaccgca	2100
tcgagctgaa	gggcatcgac	ttcaaggagg	acggcaacat	cctggggcac	aagctggagt	2160
acaactacaa	cagccacaac	gtctatatca	tgcccgacaa	gcagaagaac	ggcatcaagg	2220
tgaacttcaa	gatccgccac	aacatcgagg	acggcagcgt	gcagctcgcc	gaccactacc	2280
agcagaacac	ccccatcggc	gacggccccc	tgctgctgcc	cgacaaccac	tacctgagca	2340
cccagtcgcg	cctgagcaaa	gaccccaacg	agaagcgcg	tcacatgggt	ctgctggagt	2400
tcgtgaccgc	cgccggggtc	actctcgcca	tggaagagct	gtacaagtaa	gaattcactc	2460
ctcaggtgca	ggctgcctat	cagaagggtg	tggtctggtg	ggccaatgcc	ctgggtcaca	2520
aataccactg	agatcttttt	ccctctgcca	aaaattatgg	ggacatcatg	aagccccttg	2580
agcatctgac	ttctggctaa	taaaggaaat	ttattttcat	tgcaatagtg	tggttgaatt	2640
ttttgtgtct	ctcactcgga	aggacatatg	ggagggcata	tcatttataa	catcagaatg	2700
agtatttggg	ttagagtttg	gcaacatatg	ccatatgctg	gctgccatga	acaaaggtgg	2760
ctataaagag	gtcatcagta	tatgaaacag	ccccctgctg	tccattcctt	attccataga	2820
aaagccttga	cttgaggtta	gatttttttt	atattttggt	ttgtgttatt	ttttcttta	2880
acatccctaa	aattttcctt	acatgtttta	ctagccagat	ttttcctcct	ctcctgacta	2940
ctcccagtc	tagctgtccc	tottctctta	tgaagatccc	tcgacctgca	gccccagctt	3000
ggcgtaatac	tggtcatagc	tggttctgtg	gtgaaattgt	tatccgctca	caattccaca	3060
caacatacga	gcccgaagca	taaagtgtaa	agcctggggg	gcctaataag	tgagctaact	3120
cacattaatg	gcggttgctg	cactgcccgc	tttccagtcg	ggaaacctgt	cgtgccagcg	3180
gatccgcctc	tcaatttagt	agcaaccata	gtcccgcctc	taactccgcc	catcccgcct	3240
ctaactccgc	ccagttccgc	ccattctccg	ccccatggct	gactaatttt	ttttattttat	3300
gcagagggcg	aggccgcctc	ggcctctgag	ctattccaga	agtagtgagg	aggctttttt	3360
ggaggcctag	gcttttgcaa	aaagctaact	tggtttattg	agcttataat	ggttacaagt	3420
aaagcaatag	catcacaaat	ttcacaaata	aagcattttt	ttcactgcat	tctagtgtgt	3480
gtttgtccaa	actcatcaat	gtatcttatt	atgtctggat	ccgctgcatt	aatgaatcgg	3540
ccaacgcgcg	gggagaggcg	gtttgctgat	tgggcgctct	tcgcttccct	ggttacaagt	3600
ctcgctgcgc	tcggctcgct	ggctgcggcg	agcgggtatc	gctcactcaa	aggcggtaat	3660
acgggtattc	acagaatcag	gggataacgc	aggaaagaac	atgtgagcaa	aaggccagca	3720
aaaggccagg	aaccgtaaaa	aggccgcgtt	gctggcgctt	ttccataggg	tcggccgcct	3780
tgacgagcat	cacaaaaatc	gacgctcaag	tcagaggtgg	cgaaacccga	caggactata	3840
aagataccag	gcgtttcccc	ctgggaagct	cctcgctgcg	tctcctgttc	cgacctgccc	3900
gcttaccgga	tactttcgcc	ctttctcccc	ctcgggaagg	gtggcgcttt	ctcaatgctc	3960
acgctgtagg	tatctcagtt	cggtgtaggt	cgctcgctcc	aagctgggct	gtgtgcacga	4020
accccccggt	cagcccgacc	gctgcgcctt	atccggtaac	tatcgtcttg	agtccaacct	4080
ggtaagacac	gacttatcgc	cactggcagc	agccactggg	aacaggatta	gcagacgag	4140
glatgtaggc	ggtgctacag	agttcttgaa	gtggtggcct	aactacggct	acactagaag	4200
gacagtatct	ggtatctgcg	ctctgctgaa	gccagttacc	ttcggaaaaa	gagttggtag	4260
ctcttgatcc	ggcaaacaaa	ccaccgctgg	tagcgggtgg	ttttttgttt	gcaagcagca	4320
gattacgcgc	agaaaaaaag	gatctcaaga	agatcctttg	atctttttct	cggggtctga	4380
cgctcagttg	aacgaaaact	cacgttaagg	gattttgggt	atgagattat	caaaaaggat	4440
cttcacctag	atccttttaa	attaaaaatg	aagtttttaa	tcaatctaaa	gtatatatga	4500
gtaaaacttg	tctgacagtt	accaatgctt	aatcagtgag	gcacctatct	cagcgatctg	4560
tctatttcgt	tcatccatag	ttgcctgact	ccccgctcgt	tagataacta	cgatacggga	4620
gggcttacca	tctggcccca	gtgctgcaat	gataccgcga	gacccacgct	caccggctcc	4680
agatttatca	gcaataaacc	agccagccgg	aaggggcgag	cgcagaagtg	gtcctgcaac	4740
tttatccgcc	tccatccagt	ctattaattg	ttgcggggaa	gctagagtaa	gtagttcgcc	4800
agttaatagt	ttgcgcaacg	ttgttgccat	tgctacaggc	atcgtgggtg	cacgctcgct	4860
gtttgggtat	gcttcattca	gctccgggtc	ccaacgatca	aggcgagtta	catgatcccc	4920
catgtttgtg	aaaaaagcgg	ttagctcctt	cggtcctccg	atcgttgtca	gaagtaagtt	4980
ggccgcagtg	ttatcactca	tggttatggc	agcactgcat	aattctctta	ctgtcatgcc	5040
atccgtaaga	tgcttttctg	tgactgggtg	gtactcaacc	aagtcattct	gagaatagtg	5100
tatgcggcga	ccgagttgct	cttgcccggc	gtcaatacgg	gataataccg	cgccacatag	5160
cagaacttta	aaagtgtctc	tcattggaaa	acgttcttcg	gggcgaaaaa	tctcaaggat	5220
cttaccgctg	gttcgatcca	gttcgatgta	acccactcgt	gcacccaact	gatcttcagc	5280
atctttttact	ttcaccagcg	tttctgggtg	agcaaaaaca	ggaaggcaaa	atgccgcaaa	5340
aaaggggaata	agggcgacac	ggaaatgttg	aatactcata	ctcttccttt	ttcaatatta	5400
ttgaagcatt	tatcagggtt	attgtctcat	gagcggtatac	atatttgaat	gtatttagaa	5460
aaataaacia	ataggggttc	cgcgcacatt	tccccgaaaa	gtgccacctg		5510

<210> 72

<211> 282

<212> DNA

<213> Artificial Sequence

-39-

<220>
<223> attp

<400> 72
ccttgcgcta atgctctggt acaggtcact aataccatct aagtagttga ttcatagtga 60
ctgcatatgt tgtgtttttac agtattatgt agtctgtttt ttatgcaaaa tctaatttaa 120
tatattgata tttatatcat tttacgtttc tcgttcagct tttttatact aagttggcat 180
tataaaaaag cattgcttat caatttggtg caacgaacag gtcactatca gtcaaaataa 240
aatcattatt tgatttcaat tttgtccac tccctgcctc tg 282

<210> 73
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 73
ggccccgtaa tgcagaagaa 20

<210> 74
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 74
ggtttaaagt gcgctcctcc aagaacgtca tc 32

<210> 75
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 75
agatctagag ccgccgctac aggaacaggt ggtggcggcc 40

<210> 76
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 5PacSV40

<400> 76
ctgttaatta actgtggaat gtgtgtcagt taggggtg 37

<210> 77
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer Antisense Zeo

<400> 77
tgaacagggt cacgtcgtcc 20

<210> 78
<211> 24

-40-

<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 5' HETS

<400> 78
gggccgaaac gatctcaacc tatt 24

<210> 79
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 3' HETS

<400> 79
cgcagcggcc ctctactc 19

<210> 80
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 5BSD

<400> 80
accatgaaaa catttaacat ttctcaaca 29

<210> 81
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer SV40polyA

<400> 81
tttatttgtg aaatttgtga tgctattgc 29

<210> 82
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 3BSP

<400> 82
ttaatttcgg gtatatttga gtgga 25

<210> 83
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer EPO5XBA

<400> 83
tatctagaat ggggggtgcac gaatgtcctg cc 32

<210> 84
<211> 32
<212> DNA

-41-

<213> Artificial Sequence

<220>

<223> Primer EPO3SBI

<400> 84

tacgtacgtc atctgtcccc tgtcctgcag gc

32

<210> 85

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer GENEP03BSI

<400> 85

cgtacgtcat ctgtcccctg tcctgca

27

<210> 86

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer GENEP05XBA

<400> 86

tctagaatgg ggggtgcacgg tgagtact

28

<210> 87

<211> 4862

<212> DNA

<213> Artificial Sequence

<220>

<223> pD2eGFP-1N plasmid from Clontech

<400> 87

tagttattaa	tagtaatcaa	ttacgggggtc	attagttcat	agcccatata	tggagttccg	60
cgttacataa	cttacggtaa	atggcccgcgc	tggctgaccg	cccaacgacc	cccggccatt	120
gacgtcaata	atgacgtatg	ttcccatagt	aacgcccaata	gggactttcc	attgacgtca	180
atgggtggag	tattttacgg	aaactgcccc	cttggcagta	catcaagtgt	atcatatgcc	240
aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atggcccagta	300
catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	tcgctattac	360
catggtgatg	cggtttttgg	agtacatcaa	tgggcggtgga	tagcgggtttg	actcacgggg	420
atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	480
ggactttcca	aaatgtcgta	acaactccgc	cccattgacg	caaatgggcg	gtaggcgtgt	540
acgggtgggag	gtctatatata	gcagagctgg	tttagtgaac	cgtcagatcc	gctagcgcta	600
ccggactcag	atctcgagct	caagcttcga	attctgcagt	cgacgggtacc	gcggggcccgg	660
gatccaccgg	tcgccaccat	ggtgagcaag	ggcgaggagc	tggtcaccgg	ggtggtgccc	720
atcctgggtcg	agctgggacgg	cgacgtaaac	ggccacaagt	tcagcgtgtc	cggcgaggggc	780
gagggcgatg	ccacctacgg	caagctgacc	ctgaagttca	tctgcaccac	cggcaagctg	840
cccgtgccct	ggcccaccct	cgtgaccacc	ctgacctacg	gcgtgcagtg	cttcagccgc	900
taccccgacc	acatgaagca	gcacgacttc	ttcaagtccg	ccatgcccgga	aggctacgtc	960
caggagcgca	ccatcttctt	caaggacgac	ggcaactaca	agaccgcgcg	cgaggtgaag	1020
ttcgagggcg	acaccctggt	gaaccgcctc	gagctgaagg	gcctcgactt	caaggaggac	1080
ggcaacatcc	tggggcacaa	gctggagtac	aactacaaca	gccacaacgt	ctatatcatg	1140
gccgacaagc	agaagaacgg	catcaagggtg	aacttcaaga	tccgccacaa	catcgaggac	1200
ggcagcggtc	agctcgccga	ccactaccag	cagaacacc	ccatcggcga	cgcccccgtg	1260
gtgctgcccg	acaaccacta	cctgagcacc	cagtcgcgcc	tgagcaaaaga	cccaacagag	1320
aagcgcgacg	acatggtcct	gctggagtgc	gtgacgcgcg	ccgggatcac	tctcggcatg	1380
gacgagctgt	acaagaagct	tagccatggc	ttcccgcggg	agggtggagga	gcaggatgat	1440
ggcagcgtgc	ccatgtcttg	tgcccaggag	atcggggatgg	accgtcaccc	tcgagcctgt	1500
gcttctgcta	ggatcaatgt	gtagatgcgc	ggccgcgact	ctagatcata	atcagccata	1560
ccacatttgt	agagggttta	cttgctttta	aaaacctccc	acacctcccc	ctgaacctga	1620
aacataaaat	gaatgcaatt	gttggtgtta	acttggttat	tgcagcttat	aatgggtaca	1680

-42-

```

aataaagcaa tagcatcaca aatttcacaa ataaagcatt tttttcactg cattctagtt 1740
gtgggtttgtc caaactcatc aatgtatctt aaggcgtaaa ttgtaagcgt taatatatttg 1800
ttaaaattcgc cgttaaattt ttgttaaato agctcatttt ttaaccaata ggccgaaatc 1860
ggcaaaatcc cttataaatc aaaagaatag accgagatag ggttgagtgt tgttccagtt 1920
tggaacaaga gtccactatt aaagaacgtg gactccaacg tcaaaggcg aaaaaccgtc 1980
tatcagggcg atggccact acgtgaacca tcaccctaata caagtttttt ggggtcgagg 2040
tgccgtaaaag cactaaatcg gaaccctaaa gggagccccc gatttagagc ttgacgggga 2100
aagccggcga acgtggcgag aaaggaaggg aagaaagcga aaggagcggg cgtagggcg 2160
ctggcaagtgt tagcggtcac gctgcgcgta accaccacac ccgccgcgct taatgcgcgc 2220
ctacagggcg cgtcaggtgg cacttttcgg ggaatgtgc gcggaacccc tatttgttta 2280
tttttctaaa tatgtatccg ctcatgagac aataaccctg ataatgctt 2340
caataatatt gaaaaaggaa gagtcccgag gcggaagaa ccagctgttg aatgtgtgtc 2400
agttaggggtg tggaaagtcc ccaggtcccc cagcaggcag aagtatgcaa agcatgcatc 2460
tcaattagtc agcaaccagg tgtggaaagt ccccaggctc cccagcaggc agaagtatgc 2520
aaagcatgca tctcaattag tcagcaacca tagtcccgcc cctaactccg cccatcccg 2580
ccctaactcc gccagttccc gcgccctctc agccccatgg ctgactaatt ttttttattt 2640
atgcagggcg cgagccgccc tccgctctcg agctattcca gaagttagtg ggaggctttt 2700
ttggaggcct aggtttttgc aaagatcgat caagagacag gatgaggatc gtttcgcatg 2760
attgaacaag atggattgca cgcaggttct cccgcccgtt ggggtggagag gctattcggc 2820
tatgactggg aacaacagac aatcggctcg ccgtgttccg cctgtcgcgc 2880
caggggcccgc cggttctttt tgtcaagacc gacctgtccg gtgccctgaa tgaactgcaa 2940
gacggggcag cgcggctatc gtggctggcc acgacgggcg ttccttgccc agctgtgtct 3000
gacgttgtca ctgaagcggg tccgtccgag ctgctattgg gcgaagtgcc ggggcaggat 3060
ctcctgtcat ctcaccttgc tccgtccgag aaagtatcca tcatggctga tgcaatgcgg 3120
cggtctcata cgcttgatcc ggctacctgc ccattcgacc accaagcgaa acatcgcatc 3180
gagcagacac gtactcggat ggaagccggg cttgtcgatc aggatgatct ggacgaagag 3240
catcaggggc tcgcgccagc cgaactgttc cgcaggtca aggcgagcat gcccgacggc 3300
gaggatctcg tcgtgaccca tggcgatgcc tgcttgccga atatcatggt ggaaaatggc 3360
cgcttttctg gattcatcga ctgtggccgg ctgggtgtgg cggaccgcta tcaggacata 3420
gcgttggtct cccgtgatat tgctgaagag cttggcgccg aatgggctga ccgcttcttc 3480
gtgctttacg gtatcgccgc tcccgattcg cagcgcacgc ccttctatcg ccttcttgac 3540
gagttcttct gagcgggact ctgggggttcg aaatgaccga ccaagcgacg cccaacctgc 3600
catcacgaga ttctgattcc accgcccgtt tctatgaaag gttgggcttc ggaatcgttt 3660
tccgggacgc cggctggatg atcctccagc gcggggatct catgctggag ttcttcgccc 3720
accctagggg gaggctaact gaaacacgga agggagacaat accggaagga acccgcgcta 3780
tgacggcaat aaaaagacag acgggtgttg ctcgtttggt cataaacgcg 3840
gggttcgggtc ccagggtcgg cactctgtcg atacccacc gagaccccat tggggccaat 3900
acgcccgcgt ttcttctctt tccccacccc accccccaag ttcgggtgaa ggcccagggc 3960
tcgcagccaa cgtcggggcg gcaggccctg cctagccctc aggttactca tatatacttt 4020
agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata 4080
atctcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccccgtag 4140
aaaagatcaa aggatcttct tgagatcctt ttttctgccc cgtaatctgc tgcttgcaaa 4200
caaaaaaac accgctacca gcggtggttt gtttgccgga tcaagagcta ccaactcttt 4260
ttccgaaggt aactggcttc agcagagcgc agataccaaa tactgtcctt ctagtgtagc 4320
cgtagttagg ccaccacttc aagaactctg tagcaccgcc tacataacct cctctgctaa 4380
tcctgttacc agtggtcgt gccagtgccg ggcagcggg tcttaccggg ttggactcaa 4440
gacgatagtt accggataag gcgcagcggg tgagatacct acagcgtgag ctatgagaaa 4500
ccagcttggg gcgaacgacc tacaccgaac acagcgtgag ctatgagaaa 4560
gcgccacgct tcccgaaggg agaaaggcgg acaggtatcc ggttaagcggc agggctcgga 4620
caggagagcg cagaggggag cttccagggg gaaacgcctg gtatctttat agtccgtgcg 4680
ggtttcgcca cctctgactt ttttgtgatg ctcgtcaggg ggccttttgc tggccttttg 4740
tatggaaaaa cgccagcaac gcggcctttt tacggttctt attctgtgga taaccgtatt 4800
ctcacatggt ctttctctgcg ttatcccctg attctgtgga taaccgtatt 4862
at

```

<210> 88

<211> 5192

<212> DNA

<213> Artificial Sequence

<220>

<223> pIRESpuro2 plasmid from Clontech

<400> 88

```

gacggatcgg gagatctccc gatccccctat ggtcgactct cagtacaate tgctctgatg 60
ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gctagtgcg 120
cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc 180

```

-43-

ttaggggtag	gcgttttgcg	ctgcttcgcg	atgtacgggc	cagatatacg	cggtgacatt	240
gattattgac	tagttattaa	tagtaataca	ttacgggggc	attagttcat	agcccatata	300
tggagttccg	cgttacataa	cttacggtaa	atggcccgcc	tggctgaccg	cccaacgacc	360
cccgcccat	gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	420
attgacgtca	atgggtggac	tatttacggg	aaactgcccc	cttggcagta	catcaagtgt	480
atcatatgcc	aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	540
atgcccagta	catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	600
tgcctattac	catggtgatg	cggttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	660
actcacgggg	atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	720
aaaatcaacg	ggactttcca	aaatgtcgta	acaactccgc	cccattgacg	caaatgggcg	780
gtaggcgtgt	acgggtgggag	gtctatataa	gcagagctct	ctgggctaact	agagaaccca	840
ctgcttactg	gcttatcgaa	attaatacga	ctcactatag	ggagacccaa	gcttgggtacc	900
gagctcggat	cgatatctgc	ggcctagcta	gcgcttaagg	cctgttaacc	ggtcgtacgt	960
ctccggattc	gaattcggat	ccgcggccgc	atagataaact	gatccagtg	gctgggaatta	1020
atcgcttgag	tgcgagggcc	agctgttggg	gtgagtactc	cctctcaaaa	gcgggcatga	1080
cttctgcgct	aagattgtca	gtttccaaaa	acgaggagga	tttgatattc	acctggcccg	1140
cggtgatgcc	tttgagggtg	gccgcgtcca	tctgggtcaga	aaagacaatc	tttttgttgt	1200
caagcttgag	cttgaggcagg	cttgagatct	gtgcatacac	ttgagtgaac	atgacatcga	1260
ctttgcectt	ctctccacag	gtgtccactc	ccaggtccaa	ctgcaggtcg	agcatgcac	1320
tagggcggcc	aattccgccc	ctctccctcc	ccccccctta	acgttactgg	ccgaagccgc	1380
ttggaataag	gccggtgtgc	gtttgtctat	atgtgatttt	ccaccatatt	gccgtctttt	1440
ggcaatgtga	gggcccggaa	acctggccct	gtcttcttga	cgagcattcc	taggggtctt	1500
tcccctctcg	ccaaaggaat	gcaaggtctg	ttgaatgtcg	tgaagggaagc	agttcctctg	1560
gaagcttctt	gaagacaaac	aacgtctgta	gcgacccttt	gcaggcagcg	gaacccccca	1620
cctggcgaca	ggtgcctctg	cggccaaaag	ccacgtgtat	aagatacacc	tgcaaaaggcg	1680
gcacaacccc	agtgccacgt	tgtgagttgg	atagttgtgg	aaagagtcaa	atgggtctcc	1740
tcaagcgtat	tcaacaaggg	gctgaaggat	gccagaaagg	taccccatgt	tatgggtact	1800
gatctggggc	ctcgggtgcac	atgctttaca	tgtgtttagt	cgagggttaa	aaaacgtcta	1860
ggccccccga	accacgggga	cgtggttttc	ctttgaaaaa	cacgatgata	agcttgccac	1920
aaccacacac	gagacgacct	tccatgaccg	atgacaagcc	cacggtgcgc	ctcgccaccc	1980
gcgacgacgt	cccccgggcc	gtacgcaccc	tgccgcgcgc	gttcgcccgc	taccccgcca	2040
cgcgccacac	cgctcgaccg	gaccgccaca	tcgagcgggt	caccgagctg	caagaactct	2100
tccctcacgc	cgctcgggctc	gacatcggca	aggtgtgggt	cgcgacgacg	ggcgcccgcc	2160
tggcggtctg	gaccacggcg	gagagcgtcg	aagcgggggc	ggtgttcgcc	gagatcgccc	2220
cgcgcatggc	cgagttgagc	ggttcccggc	tggcccgcca	gcaacagatg	gaaggcctcc	2280
tggcgccgca	ccggccccag	gagcccgctg	ggttctctgg	caccgtcggc	gtctcgcccg	2340
accacagggg	caagggtctg	ggcagcgccg	tcgtgtctcc	cggagtggag	gcggcgagcg	2400
gcgcgggggt	gcccgccttc	ctggagacct	ccgcgcggcg	caacctcccc	ttctacgagc	2460
ggctcggctt	caccgtcacc	gccgacgtcg	agtccccgaa	ggaccgcgcg	acctgggtga	2520
tgaccgcgaa	gcccggtgcc	tgacgcccgc	ccagcagccc	gcagcgcccg	accgaaaagg	2580
gcgcacgacc	ccatggctcc	gaccgaagcc	gaccggggcg	gccccgcccga	ccccgcaccc	2640
gcccccgagg	cccaccgact	ctagagctcg	ctgactcagc	tcgactgtgc	cttctagttg	2700
ccagccatct	gttgtttgcc	cctcccccg	gccttccctg	accctgggaag	gtgccaactc	2760
caactgtcct	tcctaataaa	atgaggaaat	tgcactcgcat	tgtctgagta	ggtgtcatte	2820
tattctgggg	ggtgggggtg	ggcaggacag	caagggggag	gattgggaag	acaatagcag	2880
gcattgctgg	gatgcgggtg	gctctatggc	ttctgaggcg	gaaagaacca	gctggggctc	2940
gagtgcatte	tagttgtggt	ttgtccaaac	tcataaatgt	atcttatcat	gtctgtatac	3000
cgctgacctc	tagctagagc	ttggcgtaat	catggtcata	gctgtttcct	gtgtgaaatt	3060
gttatccgct	cacaattcca	cacaacatac	gagccgggaag	cataaaagtgt	aaagcctggg	3120
gtgcctaata	agtgagctaa	ctcacattaa	ttgcgttgcg	ctcactgccc	gctttccagt	3180
cgggaaacct	gtcgtgccag	ctgcattaat	gaatcgccca	acgcgcgggg	agaggcggtt	3240
tgcgtattgg	gcgctcttcc	gcttctcgcg	tcactgactc	gctgcgctcg	gtcgttcggc	3300
tgcggcgagc	ggtatcagct	cactcaaaag	cggtaatacg	ggtatccaca	gaatcagggg	3360
ataacgcagg	aaagaacatg	tgagcaaaa	gccagcaaaa	ggccaggaac	cgtaaaaagg	3420
ccgcgttgct	ggcggttttc	catagggtcc	gccccctga	cgagcatcac	aaaaaacgac	3480
gctcaagtca	gaggtggcga	aaccggacag	gactataaag	ataccaggcg	tttccccctg	3540
gaagctccct	cgtgcgctct	cctgttccga	ccctgcccgt	taccggatac	ctgtccgctc	3600
ttctcccttc	gggaagcgtg	gcgctttctc	aatgctcacg	ctgtagggtat	ctcagttcgg	3660
tgtaggctcg	tcgctccaag	ctgggctgtg	tgcacgaacc	ccccgttcag	cccgaccgct	3720
ggcccttate	cggtaaactat	cgtcttgagt	ccaacccggg	aagacacgac	ttatcgccac	3780
tggcagcagc	cactgttaac	aggattagca	gagcgaggta	tgtaggcggt	gctacagagt	3840
tcttgaagtg	gtggcctaac	tacggctaca	ctagaaggac	agtatttggt	atctgcgctc	3900
tgtgaagcc	agttaccttc	ggaaaaagag	ttggtagctc	ttgatccggc	aaacaaacca	3960
ccgctggtag	cggtgggttt	ttgttttgca	agcagcagat	tacgcgcaga	aaaaaaggat	4020
ctcaagaaga	tcctttgatc	ttttctacgg	ggtctgacgc	tcagtggaa	gaaaactcac	4080
gttaagggat	tttggctcatg	agattatcaa	aaaggatctt	cacctagatc	cttttaaat	4140
aaaaatgaag	ttttaaatca	atctaaagta	tatatagata	aacttggctc	gacagttacc	4200

-44-

```

aatgcttaat cagtgaggca cctatctcag cgatctgtct atttcgttca tccatagttg 4260
cctgaactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct ggccccagtg 4320
ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca ataaaccagc 4380
cagccggaag ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc atccagtcta 4440
ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtttg cgcaacgttg 4500
ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggatatggct tcattcagct 4560
ccggttccca acgatcaagg cgagttacat gatcccccat gttgtgcaaa aaagcgggta 4620
gctccttcgg tctccgatc gttgtcagaa gtaagtggc cgcagtgtta tcaactatgg 4680
ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc ttttctgtga 4740
ctggtgagta ctcaaccaag tcattctgag aatagtgtat gggcgacccg agttgctctt 4800
gcccggcgct aatacgggat aataccgcgc cacatgcag aactttaaaa gtgctcatca 4860
ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg agatccagtt 4920
cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc accagcgttt 4980
ctgggtgagc aaaaacagga aggcaaaatg ccgcaaaaaa ggggaataagg gcgacacgga 5040
aatgttgaat actcatactc ttcttttttc aatattattg aagcatttat caggggtatt 5100
gtctcatgag cggatacata tttgaatgta tttagaaaaa taaacaaata ggggttccgc 5160
gcacatttcc ccgaaaagtg ccacctgacg tc 5192

```

<210> 89

<211> 11182

<212> DNA

<213> Artificial Sequence

<220>

<223> pAg1 Plasmid

<400> 89

```

catgccaaacc acaggggttcc cctcgggatac aaagtacttt gatccaaccc ctccgctgct 60
atagtgcagt cggcttctga cgttcagtgc agcgtcttc tgaaaaacgac atgtcgcaca 120
agtcctaagt tacgcgacag gctgcccgcc tgcccttttc ctggcggttt cttgtcgcgt 180
gttttagtgc cataaagtag aatacttgcg actagaaccg gagacattac gccatgaaca 240
agagcgccgc cgctggcctg ctgggctatg ccgcgctcag caccgacgac caggacttga 300
ccaaccaacg ggccgaactg caccggcccg gctgcaccaa gctgttttcc gagaagatca 360
ccggcaccag gcgcgaccgc ccggagctgg ccaggatgct tgaccaccta cgccctggcg 420
acgttgtgac agtgaccagg ctgaccgcc tgggcccgag caccgcgac ctactggaca 480
ttgccgagcg catccaggag gccggcgcg gctgcgtag cctggcagag ccgtgggccc 540
acaccaccac gccggcgccg cgcatggtgt tgaccgtgtt cgccggcatt gccgagttcg 600
agcgttccct aatcatcgac cgcacccgga gccggcgcgga ggccgcgaag gcccgagggcg 660
tgaagtttgg ccccgccct accctcacc ccgcacagat cgcgcacgcc cgcgagctga 720
tcgaccagga aggcgcgacc gtgaaagagg cggctgcaat gcttggcggt catcgctcga 780
ccctgtaccg cgcacttgag cgcagcgagg aagtgcgcc caccgaggcc aggcggcgcg 840
gtgccttccg ttgaccgaga ttgaccgagg ccgacccct ggccggcgcc gagaatgaac 900
gccaagagga acaagcatga aaccgcacca ggacggccag gacgaaccgt ttttcattac 960
cgaagagatc gaggcggaga tgatcgcgcc cgggtacgtg ttcgagccgc ccgcgcacgt 1020
ctcaaccgtg cggctgcatg aaatcctggc cggtttgtct gatgccaaagc tggcggcctg 1080
gccggccagc ttggccgctg aagaaaccga gcgcgcgct ctaaaaaggt gatgtgtatt 1140
tgagtaaaac agcttgcgtc atgcggtcgc tgcgtatatg atgcgatgag taaataaaca 1200
aatacgcaag gggaaacgcat gaagggtatc gctgtactta accagaaagg cgggtcaggc 1260
aagacgacca tcgcaaccca tctagccgc gccctgcaac tcgccggggc cgatgttctg 1320
ttagtcatc ccgatcccca gggcagtgcc cgcgattggg cggccgtgcg ggaagatcaa 1380
ccgctaaccg ttgtcggcat cgaccgccc acgattgacc gcgacgtgaa ggccatcgcc 1440
cggcgcgact tcgtagtgt cgcaggagcg ccccaggcg cggacttggc tgtgtccgcg 1500
atcaaggcag ccgacttcgt gctgattccg gtgcagccaa gcccttacga catatgggcc 1560
accgcgaccc tgggtggagct ggttaagcag cgcattgagg tcacggatgg aaggctacaa 1620
gcggcctttg tcgtgtcgcg ggcgatcaaa ggcaagcgca tcggcggtga ggttgccgag 1680
gcgctggccg ggtacgagct gccattctt gagtcccgta tcacgcagcg cgtgagctac 1740
ccaggcactg ccgcgcgcgg cacaaccgtt cttgaatcag aacccgaggg cgacgctgcc 1800
cgcgaggtcc aggcgctggc cgtgaaatt aaatcaaaac tcatttgagt taatgaggta 1860
aagagaaaaat gagcaaaagc acaaacacgc taagtgcgg ccgtccgagc gcacgcagca 1920
gcaaggctgc aacgttggcc agcctggcag acacgcccag catgaagcgg gtcaactttc 1980
agttgcggcg ggaggatcac accaagctga agatgtacgc ggtacgcaa ggcaagacca 2040
ttaccgagct gctatctgaa tacatcgccg agctaccaga gtaaatgagc aaatgaataa 2100
atgagtagat gaattttagc ggctaaagga ggcggtatgg aaaatcaaga acaaccaggc 2160
accgacggcg catgtgtgga ggaacggggc gttggccagg cgtgagcgcc cgtaagcgcc 2220
tgggttgtct gccggccctg caatggcact ggaaccccca agcccgagg atcggcggtg 2280
cggtcgcaaa ccatccggcc cggtacaaat cggcgcgggc ctgggtgatg acctggtgga 2340
gaagttgaag gccgcgcagg ccgccagcg gcaacgcate gaggcagaag cacgccccgg 2400

```


tgaatcgtgg	caagcggccg	ctgatcgaat	ccgcaaagaa	tcccggcaac	cgccggcagc	2460
cggtgcgcgc	tcgattagga	agccgcacca	gggcgacgag	caaccagatt	ttttcgttcc	2520
gatgctctat	gacgtgggca	cccgcgatag	tcgcagcatc	atggacgtgg	ccgtttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	gggtgatccgc	tacgagcttc	cagacgggca	2640
cgtagagggt	tccgcagggc	cgcccgccat	ggccagtgtg	tgggattacg	acctgggtact	2700
gatggcgggt	tcccatctaa	ccgaatccat	gaaccgatac	cggaagggga	agggagacaa	2760
gcccggccgc	gtgttcctgc	cacacgttgc	ggacgtactc	aagttctgcc	ggcgagccga	2820
tggcggaag	cagaaagacg	acctggtaga	aacctgcatt	cggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cgccgcctcg	gtgacgggat	ccgagggtga	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaaacc	gggcggccgg	agtacatcga	3000
gatcgagcta	gctgattgga	tgtaccgcga	gatcacagaa	ggcaagaacc	cgacgtgtcg	3060
gacgggtcac	cccgattact	ttttgatcga	tcccgccatc	ggccgttttc	tctaccgctt	3120
ggcagccgcg	gcccagggca	aggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
gtagggcagc	gctgattgga	tcaagaagtt	ctgtttcacc	gtgcgcaagc	gtgcgggtcg	3240
aaatgacctg	ccggagtacg	atlttgaagga	ggaggcgggg	caggctggcc	cgatcctagt	3300
catgcgttac	cgcaacctga	tcgagggcga	agcatccgcc	ggttcctaata	gtacggagca	3360
gatgttaggg	caaattgccc	tagcagggga	aaagggtctc	ttcctgtgga	3420	
tagcacgtac	attgggaacc	caaagccgta	cattgggaac	cggaaccgtt	acattgggaa	3480
cccaaagccg	tacattggga	accggtcaca	catgtaagt	actgatataa	aagagaaaaa	3540
aggcgatttt	tccgcctaaa	actctttaaa	acttattaaa	actcttaaaa	cccgcctggc	3600
ttgtgcataa	ctgtctggcc	agcgcacagc	cgaagagctg	caaaaagcgc	ctaccttcg	3660
gtcgtgcgcg	tccctacgcc	ccgcgccttc	gcgctcgccct	atcgccggccg	ctggccgcctc	3720
aaaaatggct	ggcctacggc	caggcaatct	accagggcgc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	cgccgcccac	atcaaggcac	cctgcctcgc	gcgtttcggt	gatgacggtg	3840
aaaacctctg	acacatgcag	ctcccggaga	cggtcacagc	ttgtctgtaa	gcggatgccg	3900
ggagcagaca	agcccgtcag	ggcgcgctcag	cgggtgttgg	cgggtgtcgg	ggcgagacca	3960
tgaccagtcg	acgtacgat	agcggagtgt	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgcacc	atatgcgggtg	tgaataaccg	cacagatgcg	taaggagaaa	4080
ataccgcata	aggcgctctt	ccgcttcctc	gctcactgac	tcgctgcgct	cggtcgttcg	4140
gctgcggcga	cggtatcag	ctcactcaaa	ggcggttaata	cggttatcca	gatgacggtg	4200
ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	4260
ggccgcgttg	ctggcgtttt	tccataggct	ccgccccctt	gacgagcatc	acaaaaatcg	4320
acgtccaagt	cagaggtggc	gaaacccgac	aggactataa	agataccagg	cgtttcccc	4380
tggaaagctcc	ctcgtgcgct	ctcctgttcc	gacctgccc	cttaccggat	acctgtccgc	4440
ctttctccct	tcgggaagcg	tgccgctttc	tcatagctca	cgctgtaggt	atctcagttc	4500
gggtgtaggtc	gttcgctcca	agctgggtcg	tgtgcacgaa	cccccgcttc	agcccgaccg	4560
ctgcgcctta	tccggttaact	atcgtcttga	gtccaaaccg	gtaagacacg	acttatcgcc	4620
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	4680
gttcttgaag	tggtggccta	actacggcta	cactagaagg	acagtatttg	gtatctgcgc	4740
tctgctgaag	ccagttacct	tcggaaaaag	agttggtagc	tcttgatccg	gcaaaacaaac	4800
caccgctggt	agcgggtgggt	tttttgtttg	caagcagcag	attacgcgca	gaaaaaaagg	4860
atctcaagaa	gttcgcttga	tctttctctac	agggtctgac	gctcagtgga	acgaaaactc	4920
acgttaagggt	atlttgggtca	tgcattctag	gtactaaaac	aattcatcca	gtaaaatata	4980
atattttatt	ttctcccaat	caggcttgat	ccccagtaag	tcaaaaaata	gctcgacata	5040
ctgttcttcc	ccgatctcct	ccctgatcga	cgggacgcag	aaggcaatgt	cataccactt	5100
gtccgcccctg	ccgcttctcc	caagatcaat	aaagccactt	actttgcca	ctttcacaaa	5160
gatgttgctg	tctcccaggt	cgccgtggga	aaagacaagt	tccctcttcg	gcttttccgt	5220
ctttaaaaaa	tcatacagct	cgcgcggatc	tttaaatgga	gtgtcttctt	cccagtttcc	5280
gcaatccaca	tcggccagat	cggtatttcag	taagtaatcc	aattcggcta	agcggctgtc	5340
taagctatcc	gtatagggac	aatccgatata	gtcgatggag	tgaaagagcc	tgatgcactc	5400
cgcatacagc	tcgataatct	tttcagggtc	ttgttcatct	tcatactctt	ccgagcaaac	5460
gacgccatcg	gcctcactca	tgagcagatt	gctccagcca	tcatgccgtt	caaagtgacg	5520
gacctttgga	acaggcagct	ttccttccag	ccatagcatc	atgtcccttt	cccgttccac	5580
atcatagggtg	gtccctttat	accggctgtc	cgtcattttt	aaatataggt	tttctatttc	5640
tcccaccagc	ttatatacct	tagcaggaga	cattccttcc	gtatctttta	cgcagcggtta	5700
tttttcgata	agttttttca	attccgggtga	tattctcatt	ttagccattt	attatttctt	5760
tcctcttttc	tacagtattt	aaagatacce	caagaagcta	attataacaa	gacgaactcc	5820
aattcactgt	tccttgcat	ctaaaacctt	aaataaccaga	aaacagcttt	ttcaaagtgt	5880
ttttcaaagt	tggcgtataa	catagtatcg	acggagccga	ttttgaaacc	gcgggtgatca	5940
caggcagcaa	cgctctgtca	cgcttacaat	caacatgcta	ccctccgcga	gatcatccgt	6000
gtttcaaacc	cggcagctta	gttgccgttc	ttccgaatag	catcggtaac	atgagcaaac	6060
tctgcccgcct	tacaacggct	ctcccgcgtg	cgccgtcccg	gactgatggg	ctgcctgtat	6120
cgagtgggtga	ttttgtccg	agctgcccgt	cggggagctg	ttggctggct	gggtggcagga	6180
tatatgtgtg	tgtaaacaaa	ttgacgctta	gaccaactta	taacacattg	cggacgtttt	6240
taattgtactg	aattaaacgc	gaattaatte	gggggatctg	gatttttagta	ctggattttg	6300
gttttaggaa	ttagaaattt	tattgataga	agttattttac	aaatacaaat	acataactag	6360
ggtttcttat	atgctcaaca	catgagcgaa	accctatagg	aaccttaatt	cccttatctg	6420

ggaactactc	acacattatt	atggagaaac	tgcagtcaaa	tctcgggtgac	gggcaggacc	6480
ggacggggcg	gtaccggcag	gctgaagtc	agctgccaga	aacccacgtc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgcctcgtgc	6600
atgcgacacg	tccgggtcgtt	gggcagcccc	atgacagcga	ccacgctctt	gaagccctgt	6660
gcctccaggg	acttcagcag	gtgggtgtag	agcgtggagc	ccagtcccgt	ccgctgggtg	6720
cggggggaga	cgtacacggg	cgactcggcc	gtccagtcgt	aggcgttgcg	tgccttccag	6780
gggcccgcgt	aggcgatgcc	ggcgaacctc	ccgtccacct	cggcgacgag	ccagggatag	6840
cgctcccgcg	gacggacgag	gtcgtccgtc	cactcctgcg	gttccctgcg	ctcggtacgg	6900
aagttgaccc	tgcttgtctc	gatgtagtgg	ttgacgatgg	tgcagaccgc	cggcatgtcc	6960
gcctcgggtg	cacggcggat	gtcggccggg	cgctcgttct	ggctcatggt	agactcgaga	7020
gagatagatt	tgtagagaga	gactgggtgat	ttcagcgtgt	cctctccaaa	tgaatgaac	7080
ttccttatat	agaggaaggt	cttgcggaag	atagtgggat	tgtgcgtcat	cccttacgtc	7140
agtggagata	tcacatcaat	ccacttgctt	tgaagacgtg	gttggaacgt	cttctttttc	7200
cacgatgctc	ctcgtgggtg	gggggtccatc	tttgggacca	ctgtcggcag	aggcatcttg	7260
aacgatagcc	tttccctttat	cgcaatgatg	gcatttgttag	gtgccacctt	ccctttctac	7320
tgtccctttt	atgaagtgc	agatagctgg	gcaatggaat	ccgaggaggt	ttcccgatatt	7380
taccctttgt	tgaaaagtct	caatagccct	ttgggtcttct	gagactgtat	ctttgatatt	7440
cttggagtag	acgagagtgt	cggtgctccac	catgttatca	catcaatcca	cttcttttga	7500
agacgtgggt	ggaacgtctt	ctttttccac	gatgctctct	gtgggtgggg	gtccatcttt	7560
gggaccactg	tcggcagagg	catcttgaac	gatagccttt	cttttatcgc	aatgatggca	7620
tttcttagtg	ccaccttctt	tttctactgt	cccttttgatg	aagtgcacaga	tagctgggca	7680
atggaatccg	aggaggtttc	ccgatattac	cctttgttga	aaagtctcaa	tagccctttg	7740
gtcttctgag	actgtatctt	tgatattctt	ggagtagacg	agagtgtcgt	gtcaccacat	7800
gttggcaagc	tgctctagcc	aatacgcaaa	ccgctctctc	ccgcgcgttg	gccgattcat	7860
taatgcagct	ggcaccgacg	gtttcccgac	tggaaagcgg	gcagtgcgag	caacgcaatt	7920
aatgtgagtt	agctcactca	ttaggcaccc	caggctttac	actttatgct	tccggctcgt	7980
atgttgtgtg	gaattgtgag	cggataacaa	tttcacacag	gaaacagcta	tgaccatgat	8040
tacgaattcg	agccttgact	agagggctga	cggatatacag	acatgataag	atacattgat	8100
gagtttggac	aaaccacaac	tagaatgcag	tgaaaaaaat	gctttatttg	tgaattttgt	8160
gatgtctatt	ctttatttgt	aaccattata	agctgcaata	aacaagttgg	ggtgggcgaa	8220
gaactccagc	atgagatccc	cgcgtcggag	gatcatccag	ccggcgtccc	ggaaaaacgat	8280
tccgaagccc	aacctttcat	agaaggcggc	ggttgaatcg	aaatctcgtg	gcacgtgtca	8340
gtcctgtccc	tcggccacga	agtgcacgca	gttgccggcc	gggtcgcgca	gggcgaactc	8400
cgcceccac	ggctgtctgc	cgatctcggg	catggccggc	ccggaggcgt	cccggaagtt	8460
cgtggacacg	acctccgcac	actcggcgta	cagctcgtcc	aggccgcgca	cccacaccca	8520
ggccagggtg	ttgtccggca	ccacctggtc	ctggaccggc	ctgatgaaca	gggtcacgtc	8580
gtcccggacc	acacccggga	agtgcgtcct	cacgaagtcc	cgggagaacc	cgagccgggtc	8640
gggtccagaac	tgcaccgctc	cggcgacgtc	gcgcgcgggt	agcaccggaa	cggcactggt	8700
caacttggcc	atgtgaccag	atttcgtca	agttgata	aaaaagcagg	ttcaatctt	8760
gcaggaattc	gatcgacact	ctcgtctact	ccaagaatat	caaagataca	gtctcagaag	8820
accaaagggc	tattgagact	tttcaacaaa	gggtaatatc	gggaaacctc	ctcggattcc	8880
attgccagc	tatctgtcac	ttcatcaaaa	ggacagtaga	aaaggaaggt	ggcacctaca	8940
aatgcccata	ttgcgataaa	ggaaaggcta	tcgttcaaga	tgcctctgcc	gacagttggtc	9000
ccaaagatgg	acccccaccc	acgaggagca	tccgtggaaaa	agaagacgtt	ccaaccacgt	9060
cttcaaagca	agtggattga	tgtgataaca	tggtggagca	cgacactctc	gtctactcca	9120
agaatatcaa	agatacagtc	tcagaagacc	aaagggctat	tgagactttt	caacaaaggg	9180
taatatcggg	aaacctcctc	ggattccatt	gccagctat	ctgtcacttc	atcaaaagga	9240
cagtagaaaa	ggaaggtggc	acctacaaat	gccatcattg	cgataaagga	aaggctatcg	9300
ttcaagatgc	ctctgccgac	agtggctcca	aagatggacc	cccacccacg	aggagcatcg	9360
tggaaaaaga	agacgttcca	accacgtctt	caaagcaagt	ggattgatgt	gatattctcca	9420
ctgacgtaag	ggatgacgca	caatcccact	atccttcgca	agaccttctt	ctatataagg	9480
aagttcattt	catttggaga	ggacacgctg	aaatcaccag	tctctctcta	caaatctatc	9540
tctctcgagc	tttcgcagat	ccgggggggg	aatgagatat	gaaaaagcct	gaactcaccg	9600
cgacgtctgt	cgagaagttt	ctgatcgaaa	agttcgacag	cgtctccgac	ctgatgcagc	9660
tctcggaggg	cgaagaatct	cgtgctttca	gcttcgatgt	aggagggcgt	ggatatgtcc	9720
tgcgggtaaa	tagctgcgcc	gatggtttct	acaaagatcg	ttatgtttat	cggcactttg	9780
catcggccgc	gctcccgtat	ccgggaagtgc	ttgacattgg	ggagttagc	gagagcctga	9840
cctattgcat	ctcccgccgt	gcacagggtg	tcacgttgca	agacctgcct	gaaaccgaac	9900
tgcccgcgtg	tctacaaccg	gtcgcggagg	ctatggatgc	gatcgctgcg	gccgatctta	9960
gccagacgag	cgggttcggg	ccattcggag	cgcaaggaaat	cggtaataac	actacatggc	10020
gtgatttcat	attgcgcgatt	gtgatcccc	atgtgtatca	ctggcaaaact	gtgatggagc	10080
acaccgtcag	tgcgtccgtc	gcgcaggctc	tcgatgagct	gatgcttttg	gccgaggact	10140
gccccgaagt	ccggcacctc	gtgcacgcgg	atctcggctc	caacaatgtc	ctgacggaca	10200
atggccgcat	aacagcggtc	attgactgga	cggagcgcat	gttcggggat	tcccaatagc	10260
aggtcgccaa	catcttcttc	tggaggccgt	ggttggcttg	tatggagcag	cagacgcgct	10320
acttcgagcg	gaggcatccg	gagcttgcat	gatcgccacg	actccgggcg	tatatgctcc	10380
gcattgggtct	tgaccaactc	tatcagagct	tgggtgacgg	caatttcgat	gatgcagctt	10440

-47-

ggggcgcagg	tcgatgacgac	gcaatcgctcc	gatccggagc	cgggactgtc	gggcgtagac	10500
aaatcgcccc	cagaagcgcg	gccgtctgga	ccgatggctg	tgtagaagta	ctcgccgata	10560
gtggaaaccg	acgccccagc	actcgctccga	gggcaaagaa	atagagtaga	tgccgacccg	10620
atctgtcgat	cgacaagctc	gagttttctcc	ataataatgt	gtgagtagtt	cccagataag	10680
ggaattagg	ttcctatagg	gtttcgctca	tgtgttgagc	atataagaaa	cccttagtat	10740
gtatttgtat	ttgtaaaata	cttctatcaa	taaaatttct	aattcctaaa	acaaaaatcc	10800
agtactaaaa	tccagatccc	ccgaattaat	tcggcggttaa	ttcagatcaa	gcttggcact	10860
ggccgctcgt	ttacaacgct	gtgactggga	aaaccctggc	gttaccacaac	ttaatcgctt	10920
tgccagcacat	ccccctttcg	ccagctggcg	taatagcgaa	gaggcccgca	ccgatcgccc	10980
ttcccaacag	ttgcgcagcc	tgaatggcga	atgctagagc	agcttgagct	tggatcagat	11040
tgtcgtttcc	cgccttcagt	ttaaactatc	agtgtttgac	aggatatatt	ggcgggtaaa	11100
cctaagagaa	aagagcggtt	attagaataa	cggatatatta	aaaggcgctg	aaaagggtta	11160
tccgttcgct	catttgtatg	tg				11182

<210> 90

<211> 8428

<212> DNA

<213> Artificial Sequence

<220>

<223> pCambia3300 Plasmid

<400> 90

catgccaacc	acagggttcc	cctcgggata	aaagtacttt	gatccaaccc	ctccgctgct	60
atagtgcagt	cggcttctga	cggttcagtgc	agccgtcttc	tgaaaacgac	atgtcgcaca	120
agtcctaagt	tacgcgacag	gctgcgcgcc	tgcccttttc	ctggcggttt	cttgtcgctg	180
gttttagtgc	cataaagtag	aatacttgcg	actagaaccg	gagacattac	gccatgaaca	240
agagcgccgc	cgctggcctg	ctgggctatg	cccgctcag	caccgacgac	caggacttga	300
ccaaccaacg	ggccgaactg	cacgcggcgg	gctgcaccaa	gctgttttcc	gagaagatca	360
ccggcaccag	gcgcgaccgc	ccggagctgg	ccaggatgct	tgaccaccta	cgccctggcg	420
acgttgtgac	agtgaccagg	ctagaccgcc	tgcccgcgag	caccgcgcag	ctactggaca	480
ttgcgcagcg	catccaggag	gcccggcgcg	gcctgcgtag	cctggcagag	ccgtgggccc	540
acaccaccac	ccggcccgcc	cgcattggtg	tgaccgtggt	cgccggcatt	gccgagttcg	600
agcgttccct	aatcatcgac	cgcaccgcga	gcgggcgcg	ggccgccaag	gcccagaggcg	660
tgaagtttgg	ccccgccttc	accctcaccc	cggcacagat	cgcgcacgcc	cgcgagctga	720
tcgaccagga	aggccgcacc	gtgaaagagg	cggctgcact	gcttggcgctg	catcgctcga	780
ccctgtaccg	cgcacttgag	cgcagcgagg	aagtgcagcc	caccgaggcc	aggcgggcg	840
gtgccttccg	tgaggacgca	ttgaccgagg	cgacgccct	ggcgcccgcc	gagaatgaac	900
gccaagagga	acaagcatga	aaccgcacca	ggacggccag	gacgaaccgt	ttttcattac	960
cgaagagatc	gaggcgggaga	tgatcgcgcc	cgggtacgtg	ttcgagccgc	ccgcgcacgt	1020
ctcaaccgtg	cggctgcatg	aaatcctggc	cggtttgtct	gatgccaagc	tgccggcctg	1080
gccgaccgac	ttggccgctg	aagaaaccga	cgcgcgccgt	ctaaaaaggt	gatgtgtatt	1140
tgagtaaaac	agctttgcgtc	atgcggctgc	tgcttatatg	atgcgatgag	taaataaaca	1200
aatacgcaag	gggaacgcat	gaaggttate	gctgtactta	accagaaagg	cgggtcaggc	1260
aagacgacca	tcgcaaccga	ctagcccgcc	gccctgcaac	tcgcgcgggg	cgatgttctg	1320
ttagtgcatt	ccgatcccca	gggcagtgcc	cgcgattggg	cggccgtgcg	ggaagatcaa	1380
ccgctaaccg	ttgtcgccat	cgcacgcgcc	acgattgacc	gcgacgtgaa	ggccatcgcc	1440
cggcgcgact	tcgtagtgat	cgcaggagcg	cccaggcg	cggacttggc	tgtgtccgct	1500
atcaaggcag	cgcacttcgt	gctgattccg	gtgcagccaa	gcccttacga	catatggggc	1560
accgcgcacc	tggtggagct	ggttaagcag	cgcattgagg	tcacggatgg	aaggctacaa	1620
gcggcctttg	tcgtgtcgcg	ggcgatcaaa	ggcacgcgca	tcggcggtga	ggttgccgag	1680
gcgctggcgg	ggtacgagct	gcccattctt	gagtcgccga	tcacgcagcg	cgtgagctac	1740
ccaggcactg	ccgcgcgcgg	cacaaccggt	cttgaatcag	aaccgcaggg	cgcgctgccc	1800
cgcgaggtcc	aggcgctggc	cgtgaaatt	aaatcaaaac	tcatttgagt	taatgaggta	1860
aagagaaaat	gagcaaaaag	acaaacacgc	taagtggcgg	ccgtccgagc	gcacgcagca	1920
gcaaggctgc	aacgttggcc	agcctggcag	acacgccagc	catgaagcgg	gtcaactttc	1980
agttgcgcgg	ggaggatcac	accaagctga	agatgtacgc	ggtacgcca	ggcaagacca	2040
ttaccgagct	gctatctgaa	tacatcgcg	agctaccaga	gtaaaatgag	aaatgaataa	2100
atgagtagat	gaatttttag	ggctaaagga	ggcgcatgg	aaaatcaaga	acaaccaggc	2160
accgacgcgg	tggaaatgcc	catgtgtgga	ggaacgggcg	gttggccagg	cgtaagcgcc	2220
tgggttgtct	gccggccctg	caatggcact	ggaaccccca	agcccgagg	atcgccgtga	2280
cggctcgcaaa	ccatccggcc	cggtaacaa	cggcgcgccg	ctgggtgatg	acctgggtga	2340
gaagttgaag	gccgcgcagg	ccgcccagcg	gcaacgcac	gaggcagaag	cacgcccagg	2400
tgaatcgtgg	caagcgcccg	ctgatcgat	ccgcaagaa	tcccggaac	cgcccgccgc	2460
cgggtgcgcgg	tcgattagga	agccgcccga	ggcgacagag	caaccagatt	ttttcgttcc	2520
gatgctctat	gacgtgggga	cccgcgatag	tcgcagcatc	atggacgtgg	ccgttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	ggtgatccgc	tacgagcttc	cagacgggca	2640

-48-

cgtagagggt	tccgcagggc	cgcccgccat	ggccagtggt	tgggattacg	acctgggtact	2700
gatggcggtt	tcccattctaa	cogaatccat	gaaccgatac	cggggaaggga	agggagacaa	2760
gcccggccgc	gtgttccgtc	cacacgttgc	ggacgtactc	aagttctgcc	ggcgagccga	2820
tggcggaag	cagaaagacg	acctggtaga	aacctgcatt	cggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cgcccgctcg	gtgacggtat	ccgagggtga	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaaacc	ggcgggccgg	agtacatcga	3000
gatcgagcta	gctgattgga	tgtaccgcga	gatcacagaa	ggcaagaacc	cggacgtgct	3060
gacggttcac	cccgattact	ttttgatcga	tcccgcatc	ggccggtttc	tctaccgctt	3120
ggcacgccgc	gcccaggca	aggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
cagtggcagc	gcccggagag	tcaagaagtt	ctggttcacc	gtgcgcaagc	tgatcgggtc	3240
aaatgacctg	ccggagtagc	atttgaagga	ggaggcgggg	caggctggcc	cgatcctagt	3300
catgcgctac	cgcaacctga	tcgagggcga	agcatccgcc	ggttcctaata	gtacggagca	3360
gatgctaggg	caaattgccc	tagcagggga	aaaaggtcga	aaaggtctct	ttcctgtgga	3420
tagcacgtac	attgggaacc	caaagccgta	catctgggaac	cggaaccggt	acattgggaa	3480
cccaaagccg	tacattggga	accggtcaca	catgtaagtg	actgatataa	aagagaaaaa	3540
aggcgatttt	ctgcctaaaa	actctttaaa	acttattaaa	actcttaaaa	cccgcctggc	3600
ctgtgcataa	tctgtggccc	agcgacacgc	cgaagagctg	caaaaagcgc	ctaccttcgc	3660
gtcgctgcgc	tccctacgcc	ccgcgcgttc	gcgtcgccct	atcgccggccg	ctggccgcctc	3720
aaaaattggt	ggcctacggc	caggcaatct	accaggcgcc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	agcgccccc	atcaaggcac	cctgcctcgc	gcgtttcggg	gatcagcggt	3840
aaaacctctg	acacatgcag	ctcccgagga	cggtcacagc	ttgtctgtaa	gccgatgccg	3900
ggagcagaca	agcccgctcag	ggcgcgctcag	cggtgtgtgg	cggtgtgtcg	ggcgagacca	3960
tgaccagctc	acgtagcgat	agcggaggtg	atactggcct	aactatgcgg	catcagagca	4020
gattgtactg	agatgacacc	atatgcgggtg	tgaataaccg	cacagatgcg	taaggagaaa	4080
ataccgcctc	aggcgctctt	ccgcttcctc	gctcactgac	tcgctgcgct	cggtcgttcg	4140
gctgcggcga	gcggtatcag	ctcactcaaa	ggcggttaata	cggttatcca	cagaatcagg	4200
ggataacgca	gaaagaacaa	tgtgagcaaa	agggccagca	aaggccagga	accgtaaaaa	4260
ggccgcggtg	ctggcggtttt	tccatagggt	ccgccccctt	gacgagcatc	acaaaaatcg	4320
acgctcaagt	cagaggtggc	gaaacccgac	aggactataa	agataccagg	cggtttcccc	4380
tggaagctcc	ctcgctgcgt	ctcctgttcc	gacctgtccg	cttaccggat	acctgtccgc	4440
ctttctccct	tcgggaagcg	tggecgcttc	tcatagctca	cgctgtaggt	atctcagttc	4500
ggtgtagggt	gttcgctcca	agctgggctg	tgtgcacgaa	cccccgcttc	agcccgaccg	4560
ctgtctgaag	tccggttaact	atcgctctga	gtccaaccgc	gtaagacacg	acttatcgcc	4620
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	4680
gttcttgaag	tggtggccta	actacggcta	caactagaagg	acagtatttg	gtatctgcgc	4740
tctgtgaag	ccagttacct	tcggaaaaag	agttggtagc	tcttgatccg	gcaaacacaa	4800
caccgctggt	agcgggtggt	tttttgtttg	caagcagcag	attacgcgca	gaaaaaaagg	4860
atctcaagaa	gatcctttga	tcttttctac	ggggtctgac	gctcagtgga	acgaaaaactc	4920
acgttaaggg	attttgttca	tgcaataaaa	gtactaaaaa	aattcatcca	gtaaaatata	4980
atattttatt	ttctcccaat	caggcttgat	ccccagtaag	tcaaaaaata	gctcgacata	5040
ctgttcttcc	ccgatatact	ccctgatcga	ccggagcagc	aaggcaatgt	cataccactt	5100
gtccgcctcg	ccgattctcc	caagatcaat	aaagccactt	actttgccat	ctttcacaaa	5160
gatgttgctg	tctcccaggt	cgccgtggga	aaagacaagt	tctcttcggg	gcttttccgt	5220
ctttaaaaaa	tcatacagct	cgccgggac	tttaaatgga	gtgtcttctt	cccagttttc	5280
gcaatccaca	tcggccagat	cgttattcag	aattcagcca	aattcggcta	agcggctgtc	5340
taagctatct	gtatagggac	aatccgatat	gtcgatggag	tgaaagagcc	tgatgcactc	5400
cgcatacagc	tcgataatct	tttcagggtc	ttgttcatct	tcatactctt	ccgagcaaaag	5460
gacgccatcg	gcctcactca	tgagcagatt	gctccagcca	tcattgccgtt	caaagtgcag	5520
gacctttgga	acaggcagct	ttccttcag	ccatagcatc	atgtcctttt	cccgttccac	5580
atcatagggtg	gtccctttat	accggctgtc	cgctcatttt	aaatataggt	tttcattttc	5640
tcccaccagc	ttatatacct	tagcaggaga	cattccttcc	gtatctttta	cgcagcggtg	5700
tttttcgatac	agttttttca	attccgggtg	tattctcatt	ttagccattt	attattttct	5760
tccctctttt	tacagtattt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
aattcactgt	tccttgcat	ctaaaacctt	aaataccaga	aaacagcttt	ttcaaagtgt	5880
ttttcaaagt	tggcgtataa	catagtatcg	acggagccga	ttttgaaacc	gcggtgatca	5940
caggcagcaa	cgctctgtca	tcgttacaat	caacatgcta	ccctccgcga	gatcaccggt	6000
gtttcaaacc	cggcagctta	gttgccgttc	ttccgaatag	catcggtaac	atgagcaaaag	6060
tctgccgcct	tacaacggct	ctcccgctga	cgccgtcccg	gactgatggg	ctgctgtgat	6120
cgagtggtga	ttttgtgccg	agctgccggt	cggggagctg	ttggctggct	gggtggcagga	6180
tatattgtgg	tgtaacacaa	ttgacgctta	gacaacttaa	taaacacattg	cggaagcttt	6240
taatgtactg	aattaacgcc	gaattaatct	gggggatctg	gatttttagta	ctggattttg	6300
gttttaggaa	ttagaaattt	tattgataga	agttattttac	aaatacaaat	acataactaag	6360
ggtttcttat	atgtccaaca	catgagcgaa	accctatagg	aacctaat	cccttatctg	6420
ggaactactc	acacattatt	atggagaac	tcgagtcaaa	tctcggtgac	gggcaggacc	6480
ggacggggcg	gtaccggcag	gctgaagtc	agctccgaca	aaccacgcgc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgccctgtgc	6600
atgcgcacgc	tcgggtcggt	gggcagcccg	atgacagcga	ccacgctctt	gaagccctgt	6660

-49-

gcctccaggg	acttcagcag	gtgggtgtag	agcgtggagc	ccagtcccgt	ccgctgggtg	6720
cggggggaga	cgtacacggg	cgactcggcc	gtccagtcgt	aggcgttgcg	tgccctccag	6780
ggggccgcgt	aggcgatgcc	ggcgacctcg	ccgtccacct	cggcgacgag	ccagggatag	6840
cgctcccgca	gacggacgag	gtcgtccgtc	cactccctgcg	gttcctgcgg	ctcgggtacgg	6900
aagttgaccg	tgcttgctctc	gatgtagtgg	ttgacgatgg	tgacagaccgc	cgcatgtcc	6960
gcctcggtgg	cacggcggat	gtcggccggg	cgtcgttctg	ggctcatggg	agactcgaga	7020
gagatagatt	tgtagagaga	gactggtgat	ttcagcgtgt	cctctccaaa	tgaaatgaac	7080
ttccttatat	agaggaagg	cttgcggaagg	atagtgggat	tgtgcgtcat	cccttacgtc	7140
agtggagata	tcacatcaat	ccacttgctt	tgaagacgtg	gttggaaagt	cttctttttc	7200
cacgatgctc	ctcgtgggtg	gggggtccatc	tttgggacca	ctgtcggcag	aggcatcttg	7260
aacgatagcc	tttcccttat	cgcaatgatg	gcattttag	gtgccacctt	ccttttctac	7320
tgtccttttg	atgaagtgac	agatagctgg	gcaatggaat	ccgaggaggt	ttcccgatat	7380
taccctttgt	tgaaaagtct	caatagccct	ttggctcttct	gagactgtat	ctttgatatt	7440
cttggagtag	acgagagtgt	cgtgctccac	catgttatca	catcaatcca	cttgctttga	7500
agacgtgggt	ggaaagtctt	ctttttccac	gatgctcctc	gtgggtgggg	gtccatcttt	7560
gggaccactg	tcggcagagg	catcttgaac	gatagccttt	cctttatcgc	aatgatggca	7620
ttttagtggt	ccaccttctt	tttctactgt	cctttttagt	aagtgcacaga	tagctgggca	7680
atggaaatcg	aggaggtttc	ccgatattat	cctttgttga	aaagtctcaa	tagccctttg	7740
gtcttctgag	actgtatctt	tgatattctt	ggagttagacg	agagtgtcgt	gctccacctt	7800
gttggcaagc	tgctctagcc	aatacgcaaa	ccgcctctcc	ccgcgcgttg	gccgattcat	7860
taatgcagct	ggcacgacag	gtttcccgac	tggaaaagcgg	gcagtgcgag	caacgcaatt	7920
aatgtgagtt	agctcactca	ttaggcaccc	caggctttac	actttatget	tccggtcgt	7980
atgttgtgtg	gaattgtgag	cggataacaa	tttcacacag	gaaacagcta	tgaccatgat	8040
tacgaattcg	agctcggtag	ccggggatcc	tctagagtcg	acctgcaggc	atgcaagctt	8100
ggcactggcc	gtcgtttttac	aacgtcgtga	ctgggaaaac	cctggcgtta	cccaacttaa	8160
tcgccttgca	gcacatcccc	ctttcgccag	ctggcgtaat	agcgaagagg	cccgccacca	8220
tcgccttccc	caacagttgc	gcagcctgaa	tggcgaatgc	tagagcagct	tgagcttggg	8280
tcagattgtc	gtttcccgcc	ttcagtttaa	actatcagtg	tttgacagga	tatatggcg	8340
ggttaaaccga	agagaaaaga	gcgtttatta	gaataacgga	tattttaaag	ggcgtgaaaa	8400
ggtttatccg	ttcgtccatt	tgtatgtg				8428

<210> 91

<211> 3438

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attBZeo Plasmid

<400> 91

tcgaccctct	agtcaaggcc	ttaaagtgagt	cgtattacgg	actggccgctc	gttttacaaac	60
gtcgtgactg	ggaaaaccct	ggcggttaccc	aacttaatcg	ccttgacagca	catccccctt	120
tcgccagctg	gcgtaatagc	gaagaggccc	gcaccgatcg	cccttcccaa	cagttgcgca	180
gcctgaatgg	cgaatggcgc	ttcgcttggt	aataaagccc	gcttcggcgg	gctttttttt	240
gttaactacg	tcaggtggca	cttttcgggg	aaatgtgcgc	ggaaccctta	tttgtttatt	300
tttctaaata	cattcaaata	tgtatccgct	catgagacaa	taaccctgat	aaatgcttca	360
ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttattccctt	420
ttttgcggca	ttttgccttc	ctgtttttgc	tcacccagaa	acgctgggtga	aagtaaaaga	480
tgctgaagat	cagttgggtg	cacgagtggt	ttacatcgaa	ctggatctca	acagcggtaa	540
gatccttgag	agttttcgcc	ccgaagaacg	ttctccaatg	atgagcactt	ttaaagtctt	600
gctatgtggc	cgggtattat	cccggtttga	cgccgggcaa	gagcaactcg	gtcgcgcgtc	660
acactattct	cagaatgact	tgggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	720
tggcatgaca	gtaagagaat	tatgcagtcg	tgccataacc	atgagtgata	acactgcggc	780
caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	tgcaacaacat	840
gggggatcat	gtaactcgcc	ttgatcgttg	ggaaccggag	ctgaatgaag	ccataccaaa	900
cgacgagcgt	gacaccacga	tgccgtgtag	aatggcaaca	acgttgcgca	aactattaac	960
tggcgaacta	cttactctag	cttcccggca	acaatttaata	gactggatgg	aggcggataa	1020
agttgcagga	ccacttctgc	gctcggccct	tccggctggc	tggtttatgt	ctgataaaatc	1080
tggagccggg	gagcgtgggt	ctcgcggtat	catgacagca	ctggggccag	atggtaagcc	1140
ctccggtatc	gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	1200
acagatcgct	gagataggtg	cctcactgat	taagcattgg	taactgtcag	accaagttta	1260
ctcatatata	ctttagattg	atttaccctg	gttgataatc	agaaaagccc	caaaaacagg	1320
aagattgtat	aagcaaatat	ttaaattgta	aacgtttaata	ttttgttaaa	attcgcgtta	1380
aatttttgtt	aaatcagctc	attttttaac	caataggccg	aaatcggcaa	aatcccttat	1440
aatcaaaaag	aatagcccg	gatagggttg	agtgttgttc	cagtttggaa	caagagtcca	1500
ctataaaaag	acgtggactc	caacgtcaaa	gggcgaaaaa	ccgtctatca	gggcgatggc	1560
ccactacgtg	aacctaccac	caaatcaagt	tttttggggg	cgaggtgccg	taaagcacta	1620

-50-

aatcgggaacc ctaaaggagg cccccgattt agagcttgac ggggaaagcg aacgtggcga 1680
gaaagggaagg gaagaaagcg aaaggagcgg gcgctagggc gctggcaagt gtagcggtca 1740
cgctgcgcgt aaccaccaca cccgccgcgc ttaatgcgcc gctacagggc gcgtaaaagg 1800
atctaggtga agatcctttt tgataatctc atgacaaaaa tcccttaacg tgagttttcg 1860
ttccactgag cgtcagaccc cgtagaaaaa atcaaaggat cttcttgaga tccctttttt 1920
ctgcgcgtaa tctgctgctt gcaaacaaaa aaaccaccgc taccagcggg gggttggttg 1980
ccggatcaag agctaccaac tctttttccg aaggtaactg gcttcagcag agcgagata 2040
ccaaataactg ttcttctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca 2100
ccgctacat acctcgctct gctaactctg ttaccagtgg ctgctgccag tggcgataag 2160
tcgtgtctta ccgggttggg ctcaagacga tagttaccgg ataaggcgca gcggtcgggc 2220
tgaacggggg gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga 2280
tacctacagc gtgagctatg agaaagcgcc acgcttcccg aaggagaaaa ggccgacagg 2340
tatccggtaa gcggcagggg cggaaacagga gagcgcacga gggagcttcc aggggggaaac 2400
gctgggtatc tttatagtc tgcgggttt cgccacctct gacttgagcg tcgatttttg 2460
tgatgctcgt cagggggggcg gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg 2520
ttctggcct tttgctggcc ttttgcctcac atgtaatgtg agttagctca ctcattaggc 2580
accccaggct ttacacttta tgcttccggc tctatgttg tgtggaattg tgagcggata 2640
acaatttcac acaggaaaca gctatgacca tgattacgcc aagctacgta atacgactca 2700
ctagtggggc ccgtgcaatt gaagccgggt ggcgccaagc ttctctgcag gattgaagcc 2760
tgctttttta tactaacttg agcgaatct ggatccatgg ccaagttgac cagtgcggtt 2820
ccggtgctca ccgcgcgcga cgtcgccgga gcggtcgagt tctggaccga ccggtcggg 2880
ttctcccggg acttcgtgga ggacgacttc gccggtgtgg tccgggacga cgtgaccctg 2940
ttcatcagcg cggtcagga ccagggtgg tccggaacaa cctggcctg ggtgtgggtg 3000
cgcgccctgg acgagctgta ccgagtggtg tccgaggtcg tgtccacgaa cttccgggac 3060
gcctccgggc cggccatgac cgagatcggc gagcagccgt gggggcgggg gttcgccctg 3120
cgcgaccgag ccggcaactg cgtgcacttc gtggccgagg agcaggactg acacgtgcta 3180
cgagatttcg attccaccgc gccttctat gaaaggttgg gcttcggaat cgtttccgg 3240
gacgcccgtt ggatgatcct ccagcgcggg gatctcatgc tggagttctt cggccacccc 3300
aacttgttta ttgcagctta taatgggttac aaataaagca atagcatcac aaatttcaca 3360
aataaagcat tttttcact gcattctagt tgtggtttgt ccaaactcat caatgtatct 3420
tatcatgtct gtataccg 3438

<210> 92
<211> 10549
<212> DNA
<213> Artificial Sequence

<220>
<223> pCambia1302 Plasmid

<300>
<308> Genbank #AF234398
<309> 2000-04-24

<400> 92
catggtagat ctgactagta aaggagaaga acttttctact ggagttgtcc caattcttgt 60
tgaattagat ggtgatgtta atgggcacaa attttctgtc agtggagagg gtgaagggtga 120
tgcaacatac ggaaaactta cccttaaatt tatttgcaact actggaaaac tacctgttcc 180
gtggccaaca cttgtcacta ctttctctta tgggtgttcaa tgcttttcaa gatacccaga 240
tcatatgaag cggcagcact tcttcaagag cgcctatgcct gagggatacg tgcaggagag 300
gaccatcttc ttcaaggacg acgggaacta caagacacgt gctgaagtca agtttgaggg 360
agacaccctc gtcaacagga tctgagcttaa gggaaatcgat ttcaaggagg acggaaacat 420
cctcggccac aagttggaat acaactacaa ctcccacaaac gtatacatca tggccgacaa 480
gcaaaagaac ggcatacaaag ccaacttcaa gaccgcggc aacatcgaag acggcggtcg 540
gcaactcgct gatcattatc aacaaaatac tccaattggc gatggccctg tccctttacc 600
agacaaccat tacctgtcca cacaatctgc cctttcgaaa gatcccaacg aaaagagaga 660
ccacatggtc cttcttgagt ttgtaacagc tgctgggatt acacatggca tggatgaact 720
atacaaagct agccaccacc accaccacca cgtgtgaatt ggtgaccagc tgaatttcc 780
ccgatcgttc aaacatttgg caataaagtt tcttaagatt gaatcctgtt gccggtcttg 840
cgatgattat catataattt ctggtgaatt acgttaagca tgtaataatt aacatgta 900
gcatgacggt atttatgaga tgggttttta tgattagagt cccgcaatta tacattta 960
acgcgataga aaacaaaata tagcgcgcaa actaggataa attatcgcg gcggtgtcat 1020
ctatggttac agatcgggaa ttaactatc agtgtttgac aggatataat ggcgggtaaa 1080
cctaagagaa aagagcggtt attagaataa cggatattta aaagggcggtg aaaaggttta 1140
tccgttcgtc catttgatg tgcatgccaa ccacagggtt cccctcggg tcaaagta 1200
ttgatccaac ccctcgctg ctatagtgca gtcggcttct gacgttcagt gcagcgct 1260
tctgaaaacg acatgtcgca caagtccata gttacgcgac aggctgccgc cctgccctt 1320

tcctggcggtt	ttcttgtcgc	gtgttttagt	cgcataaagt	agaataacttg	cgactagaac	1380
cggagacatt	acgccatgaa	caagagcgcc	gccgctggcc	tgctgggcta	tgcccgcgtc	1440
agcaccgacg	accaggactt	gaccaacca	cgggcccgaac	tgacgcggcg	cggctgcacc	1500
aagctgtttt	ccgagaagat	caccggcacc	aggcgcgacc	gcccggagct	ggccaggatg	1560
cttgaccacc	tacgccctgg	cgacgttgtg	acagtgacca	ggctagaccg	cctggcccgc	1620
agcaccgcg	acctactgga	cattgccgag	cgcattccagg	aggccggcgc	gggccttgcgt	1680
agcctggcag	agccgtgggc	cgacaccacc	acgccggccg	gccgcatggt	gttgaccgtg	1740
ttcgccggca	ttgccgagtt	cgagcggttc	ctaatacatcg	accgcaccgc	gagcggggcg	1800
gagggcccca	aggcccagg	cgtgaagttt	ggcccccgcc	ctaccctcac	cccggcacag	1860
atcgcgcacg	cccgcgagct	gatcgaccag	gaaggccgca	ccgtgaaaga	ggcggctgca	1920
ctgcttggcg	tgcatcgctc	gaccctgtac	cgcgcacttg	agcgcagcga	ggaagtgcag	1980
cccaccgagg	ccaggcggcg	cggtgccctc	cgtgaggacg	cattgaccga	ggccgacgcc	2040
ctggcgggcg	ccgagaatga	acgccaaag	gaacaagcat	gaaaccgcac	caggacggcc	2100
aggacgaacc	gtttttcatt	accgaagaga	tgcaggcgga	gatgatcgcg	gccgggtacg	2160
tgttcgagcc	gccccgcac	gtctcaaccg	tgccgctgca	tgaaatcctg	gccgggttgt	2220
ctgatgccaa	gctggcgggc	tgcccgggca	gcttggccgc	tgaagaaacc	gagcgccgcc	2280
gtctaaaaag	gtgatgtgta	tttgagtaaa	acagcttgcg	tcagtgcggtc	gctgcgtata	2340
tgatgccgat	agtaaaataa	caaatacgca	aggggaacgc	atgaagggtta	tcgctgtact	2400
taaccagaaa	ggcgggtcag	gcaagacgac	catcgcaacc	catctagccc	gcgccttgca	2460
actcgccggg	gccgatgttc	tgttagtcca	ttccgatccc	cagggcagtg	cccgcgattg	2520
ggcgccgctg	gacataatga	aaccgctaac	cgttgtcgcc	atcgaccgcc	cgaccattga	2580
ccgcgacgtg	aaggccatcg	gccggcgcg	cttcgtagt	atcgacggag	cgcccagggc	2640
ggcggaactg	gctgtgtccg	cgatcaaggc	agccgacttc	gtgctgattc	cgggtgcagcc	2700
aagcccttac	gacatatggg	ccaccgcca	cctgggtgag	ctggttaagc	agcgcatgga	2760
gggtcacggat	ggaaggctac	aagcgccctt	tgctgtgtcg	cgggcgatca	aaggcacgcg	2820
catcgccggt	gaggttgccg	agggcgctgg	cgggtacgag	ctgcccattc	ttgagttccc	2880
tatcacgcag	cgcgtgagct	accaggcac	tgcccgccgc	ggcacaaccg	ttcttgaatc	2940
agaaccgcag	ggcgacgctg	cccgcgaggt	ccaggcgctg	gccgctgaaa	ttaaatcaaa	3000
actcatttga	gttaatgagg	taaagagaaa	atgagcaaaa	gcacaaacac	gctaagtgcc	3060
ggccgtccga	gcgcacgcag	cagcaaggct	gcaacgttgg	ccagcctggc	agacacgcca	3120
gccatgaagc	gggtcaactt	tcagttgccg	cgaggagatc	acaccaagct	gaagatgtac	3180
gcggtatgcc	aaggcaagac	cattaccgag	ctgctatctg	aatacatcgc	gcagctacca	3240
gagtaaatga	gcaaatgaat	aaatgagtga	aatgaattta	gcggtctaaag	gcggtggcat	3300
ggaaaatcaa	gaacaaccag	gcaccgacgc	cgtggaatgc	cccatgtgtg	gaggaacggg	3360
cggttggcca	ggcgtaagcg	gctgggttgt	ctgcgggccc	tgcaatggca	ctgggaaccc	3420
caagcccgag	gaactcggcg	gacggtcgca	aaccatccgg	cccggtaaaa	atcgccgcgg	3480
cgctgggtga	tgacctgggt	gagaagttga	aggccgcgca	ggccgcccag	cggcaacgca	3540
tcgaggcaga	agcacgcccc	ggtgaatcgt	ggcaagcgcc	cgctgatcga	atccgcagaa	3600
aatcccggca	accgcggca	gccggtgcgc	cgtcgattag	gaagccgccc	aagggcgacg	3660
agcaaccaga	ttttttcggt	ccgatgctct	atgacgtggg	caccgcgcat	agtcgcagca	3720
tcattggaagt	ggccgttttc	cgtctgtcga	agcgtgaccg	acgagctggc	gaggtgatcc	3780
gctacgagct	tcacagcggg	cacgtagagg	tttcgcagg	gccggccggc	atggccagtg	3840
tggtgggatta	cgacatggta	ctgatggcgg	tttcccatct	aaccgaatcc	atgaaccgat	3900
accgggaagg	gaaggagagc	aagcccggcc	gcgtgttccg	tcacacagct	gcggaacgtac	3960
tcaagttctg	ccggcgagcc	gatggcggaa	agcagaaaag	cgacctggta	gaaacctgca	4020
ccgtgcgcaa	caccacgcac	gttgccatgc	agcgtacgaa	gaaggccaag	aacggccgcc	4080
tggtgacggg	atccgagggg	gaagccttga	ttagccgcta	caagatcgta	aagagcgaaa	4140
ccgggcggcc	ggagtacatc	gagatcgagc	tagctgattg	gatgtaccgc	gagatcacag	4200
aaggcaagaa	cccggacgtg	ctgacgggtc	accccgatta	ctttttgatc	gatcccgcca	4260
tcggccggtt	tctctaccgc	ctggcacgcc	gcgcgcgagg	caaggcagaa	gccagatggg	4320
tggtcaagac	gatctacgaa	cgcagtgcca	gcgcgggaga	gttcaagaag	ttctgtttca	4380
ccgtgcgcaa	gctgatcggg	tcaaatgacc	tgccggagta	cgatttgaag	gaggaggcgg	4440
ggcaggctgg	cccgatccca	gtcatgcgct	accgcaacct	gatcgagggc	gaagcatccg	4500
ccggttccta	atgtacggag	cagatgctag	ggcaaatggc	cctagcaggg	gaaaaaggtc	4560
gaaaaggctc	gtttcctgtg	gatagcacgt	acattgggaa	cccaaagccg	tacattggga	4620
accggaaccc	gtacattggg	aaccctaaag	cgtacattgg	gaaccgggtc	cacatgtaag	4680
tgactgatat	aaaagagaaa	aaaggcgatt	tttcgcgcta	aaactcttta	aaacttatta	4740
aaactcttaa	aaccgcctg	gcctgtgctg	aaactgtctg	ccagegcaca	gccgaagagc	4800
tgcaaaaagc	gcctaccctt	cggctcgctgc	gctccctacg	ccccgcccgt	tcgcgtcggc	4860
ctatcgccgc	cgtctggccg	tcaaaaatgg	ctggccctac	gccaggcaat	ctaccagggc	4920
gcggacaagc	gcgcgcgtcg	ccactcgacc	gcggcgccgc	acatcaaggc	accctgcctc	4980
gcgcggttcc	gtgatgacgg	tgaaaacctc	tgacacatgc	agctcccggg	gacggtcaca	5040
gcttgtctgt	aagcggatgc	cgggagcaga	caagcccgct	agggcgcgct	agcgggtgtt	5100
ggcgggtgtc	ggggcgacgc	catgacccag	tcacgtagcg	atagcggagt	gtatactggc	5160
ttactatgct	ggcatcagag	cagattgtac	tgagagtgc	ccatatgcgg	tgtgaaatac	5220
cgcacagatg	cgtgaaggaga	aaataccgca	tcaggcgctc	ttccgcttcc	tcgctcactg	5280
actcgctgcg	ctcggtcggt	cggctgcggc	gagcgggtatc	agctcactca	aaggcggtaa	5340

tacggttatc	cacagaatca	ggggataacg	caggaaagaa	catgtgagca	aaaggccagc	5400
aaaaggccag	gaaccgtaaa	aaggccgcgt	tgctggcggt	tttccatagg	ctccgcccc	5460
ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagaggtg	gcgaaacccg	acaggactat	5520
aaagatacca	ggcgtttccc	cctggaagct	ccctcgtgcg	ctctcctgtt	ccgaccctgc	5580
cgcttaccgg	atacctgtcc	gcctttctcc	cttcgggaag	cgtggcgctt	tctcatagct	5640
cacgctgtag	gtatctcagt	tcgggtgtagg	tcgttcgctc	caagctgggc	tgtgtgcacg	5700
aacccccgt	tcagcccgac	cgctgcgcct	tatccggtaa	ctatcgtcct	gagtccaacc	5760
cggtaagaca	cgacttatcg	ccactggcag	cagccactgg	taacaggatt	agcagagcga	5820
ggatgttagg	cgggtctaca	gagttcttga	agtgggtggc	taactacggc	tacactagaa	5880
ggacagtatt	tggtatctgc	gctctgctga	agccagttac	cttcggaaaa	agagttggta	5940
gctcttgatc	cggcaaaaca	accaccgctg	gtagcgggtg	tttttttgtt	tgcaagcagc	6000
agattaacgg	cagaaaaaaa	ggatctcaag	aagatccttt	gatcttttct	acggggtctg	6060
acgctcagtg	gaacgaaaac	tcacgttaag	ggatttttgt	catgcattct	aggtactaaa	6120
acaattcatc	cagtaaaaata	taatattttta	ttttctccca	atcagggttg	atccccagta	6180
agtcaaaaaa	tagtctcgaca	tagtcttctt	ccccgataatc	ctccctgatc	gaccggacgc	6240
agaaggcaat	gtcataccac	ttgtccgccc	tgccgcttct	cccaagatca	ataaagccac	6300
ttactttgcc	atctttcaca	aagatgtttg	tgctctccag	gtcgcctggg	gaaaagacaa	6360
gttctcttcc	gggtcttttcc	gtcttttaaaa	aatcatacac	ctcgcgcgga	tctttaaaatg	6420
gagtgctctt	ttcccagttt	tcgcaatcca	catcgccag	atcgttatte	agtaagtaat	6480
ccaattcggc	taagcggctg	tctaagctat	tcgtataggg	acaatccgat	atgtcgatgg	6540
agtgaagag	cctgatgcac	tccgcataca	gctcgataat	cttttcaggg	tttcttctat	6600
cttcatactc	ttccgagcaa	aggacgccat	cggcctcact	catgagcaga	ttgctccagc	6660
catcatgccc	ttcaaaagtgc	aggacctttg	gaacaggcag	ctttccttcc	agccatagca	6720
tcattgtcct	ttcccgttcc	acatcatagg	tggttccctt	ataccggctg	tcctgtcatt	6780
ttaaatatag	gttttctatt	tctcccacca	gcttatatac	cttagcagga	gacattcctt	6840
ccgtatcttt	tacgcagcgg	tatttttctga	tcagtttttt	caattccggg	gatattctca	6900
ttttagccat	tattattttc	cttccctctt	tctacagtat	ttaaagatac	cccaagaagc	6960
taattataac	aagacgaact	ccaattcact	gttctttgca	ttctaaaacc	ttaaatacca	7020
gaaaaacagc	ttttcaaaagt	tgtttttcaaa	gttggcgtat	aacatagtat	cgacggagcc	7080
gattttgaaa	ccgcgggtgat	cacaggcagc	aacgctctgt	catcgttaca	atcaacatgc	7140
tacctccgcg	gagatcatcc	gtgttttcaaa	cccgcagcgt	tagttgcccgt	tcttccgaat	7200
agcatcggta	acatgagcaa	agtctgcccgc	cttacaacgg	ctctcccgcgt	gacgccgtcc	7260
cggactgatg	ggctgcctgt	atcgagtggg	gattttgtgc	cgagctgccc	gtcggggagc	7320
tgttggctgg	ctgggtggcag	gatatattgt	gggttaaaca	aattgacgct	tagacaactt	7380
aataacacat	tgccgacgtt	tttaattgtac	tgaattaaacg	ccgaattaat	tcgggggatc	7440
tggaattttag	tactggattt	tggttttagg	aattagaaaat	tttattgata	gaagtatttt	7500
acaaatacaa	atacatacta	agggtttctg	atatgctcaa	cacatgagcg	aaacctata	7560
ggaaccctaa	ttcccttatc	tggaacttac	tcacacatta	ttatggagaa	actcgagctt	7620
gtcgatcgac	agatccgggtc	ggcatctact	ctattttctt	gccctcggac	gagtgctggg	7680
gcgtcgggtg	ccactatcgg	cgagtacttc	tacacagcca	tcggtccaga	cggccgcgtc	7740
tctgcgggcg	atttgtgtac	gcccagacagt	cccggctccg	gatcggacga	ttgcgtcgca	7800
tcgaccctgc	gcccgaagctg	catcatcgaa	attgcccgtc	accaagctct	gatagagttg	7860
gtcaagacca	atgcgggagca	tatagcccgc	gactcgtggc	gatcctgcaa	gctccggatg	7920
cctccgctcg	aagtagecgc	tctgctgtct	catacaagcc	aaccacggcc	tccagaagaa	7980
gatgttgccg	acctcgtatt	gggaatcccc	gaacatcgcc	tcgctccagt	caatgaccgc	8040
tgttatgcgg	ccatttgtccg	tcaggacatt	gttggagccg	aaatccgcgt	gcacgaggtg	8100
ccggacttcg	gggcagtcct	cggcccaag	catcagctca	tcgagagcct	gcgcgacgga	8160
cgactgacg	gtgtcgtcca	tcacagtttg	ccagtgtac	acatggggat	cagcaatcgc	8220
gcataatgaaa	tcacgccatg	tagtgtattg	accgatttct	tgcggtccga	atgggcccga	8280
cccgcctcgt	tggttaagat	cggccgcagc	gatcgcaccc	atagcctccg	cgaccgggtg	8340
tagaacagcg	ggcagttcgg	tttcaggcag	gtcttgcaac	gtgacaccct	gtgcacggcg	8400
ggagatgcaa	taggtcaggc	tctcgctaaa	ctccccaatg	tcaagcactt	ccggaatcgg	8460
gagcgcggcc	gatgcaaagt	gccgataaac	ataacgatct	ttgtagaaac	catcggcgca	8520
gctattttacc	cgcaggacat	atccacgccc	tctacatcgc	aagctgaaag	cacgagattc	8580
ttcgcctctc	gagagctgca	tcaggtcgga	gacgctgtcg	aacttttctga	tcagaaactt	8640
ctcgacagac	gtcgcgggtga	gttcaggctt	tttcatactc	cattgcccc	cgggatctgc	8700
gaaagctcga	gagagataga	tttgtagaga	gagactgggt	atttcagcgt	gtcctctcca	8760
aatgaaatga	acttcccttat	atagaggaag	gtcttgcgaa	ggatagtggt	attgtgcgtc	8820
atcccttacg	tcagtggaga	tatcacatca	atccacttgc	tttgaagacg	tggttggaac	8880
gtcttctttt	tccacgatgc	tcctcgtggg	tggggggtcca	tctttggggac	cactgtcggc	8940
agaggcatct	tgaacgatag	cttttccctt	tggtcatttg	tggtcatttg	aggtgccacc	9000
ttccttttct	actgtccttt	tgatgaagtg	acagatagct	gggcaatgga	atccgaggag	9060
gtttcccgat	attacccttt	gttgaaaagt	ctcaatagcc	ctttgggtctt	ctgagactgt	9120
atcttttgata	ttccttgaggt	agacgagagt	gtcgtgtctcc	accatgttat	cacatcaatt	9180
cacttgcttt	gaagacgtgg	ttggaacgtc	ttctttttcc	acgatgctcc	tcgtgggtgg	9240
gggtccatct	ttgggaccac	tgtcggcaga	ggcatcttga	acgatagcct	ttcctttatc	9300
gcaatgatgg	catttgtagg	tgccaccttc	cttttctact	gtccttttga	tgaagtgaac	9360

-53-

```

gatagctggg caatggaatc cgaggagggtt tcccgatatt accctttggt gaaaagtctc 9420
aatagccctt tgggtcttctg agactgtatc tttgatattc ttggagtaga cgagagtgtc 9480
gtgctccacc atgtttggcaa gctgctctag ccaatacgca aaccgcctct ccccgcgctg 9540
tggccgattc attaatgcag ctggcacgac aggtttcccg actggaaagc gggcagtgag 9600
cgcaacgcaa ttaatgtgag ttagctcact cattagcac cccaggcttt acactttatg 9660
cttcgggtc gtatgttgtg tgggaattgtg agcggataac aatttcacac aggaaacagc 9720
tatgaccatg attacgaatt cgagctcggg acccggggat cctctagagt cgacctgcag 9780
gcatgcaagc ttggcactgg ccgtcggttt acaacgtcgt gactgggaaa accctggcgt 9840
tacccaactt aatcgccctg cagcacatcc ccctttcgcc agctggcgta atagcgaaga 9900
ggcccgccac gatcgccctt cccaacagtt ggcgcagcctg aatggcgaaat gctagagcag 9960
cttgagcttg gatcagattg tcgtttcccg ccttcagttt agcttcatgg agtcaaagat 10020
tcaaatagag gacctaacag aactcgccgt aaagactggc gaacagttca tacagagtct 10080
cttacgactc aatgacaaga agaaaatctt cgtcaacatg gtggagcacg acacacttgt 10140
ctactccaaa aatatcaaaag atacagtctc agaagaccaa agggcaattg agacttttca 10200
acaaagggta atatccggaa acctcctcgg attccattgc ccagctatct gtcactttat 10260
tgtgaagata gtgaaaaagg aagggtggctc ctacaaatgc catcattgag ataaaggaaa 10320
ggccatcggt gaagatgcct ctgccgacag tggtcccaaa gatggacccc caccacagag 10380
gagcatcggt gaaaaagaag acgttccaac cacgtcttca aagcaagtgg attgatgtga 10440
tatctccact gacgtaaggg atgacgcaca atcccactat ccttcgcaag acccttcctc 10500
tatataagga agttcatttc atttgagag aacacggggg actcttgac 10549

```

<210> 93
 <211> 33
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> CaMV35SpolyA Primer

<400> 93
 ctgaattaac gccgaattaa ttccgggggat ctg 33

<210> 94
 <211> 29
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> CaMV35Spr Primer

<400> 94
 ctagagcagc ttgccaacat ggtggagca 29

<210> 95
 <211> 12592
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pAg2 Plasmid

```

<400> 95
gtacgaagaa ggccaagaac ggccgcctgg tgacgggtatc cgaggggtgaa gccttgatta 60
gccgctacaa gatcgtaaaag agcgaaaaccg ggcggccgga gtacatcgag atcgagctag 120
ctgattggat gtaccgcgag atcacagaag gcaagaaccc ggacgtgctg acggttcacc 180
ccgattactt tttgatcgat cccggcatcg gccgttttct ctaccgcctg gcacgcccgcg 240
ccgcaggcaa ggcagaagcc agatggttgt tcaagacgat ctacgaacgc agtggcagcg 300
ccggagagtt caagaagttc tgtttcaccc tgccgaagct gatcgggtca aatgacctgc 360
cggagtacga tttgaaggag gaggcggggc aggtctggccc gatcctagtc atgcgctacc 420
gcaacctgat cgagggcgaa gcatccgccc gttcctaattg tacggagcag atgctagggc 480
aaattgccct agcaggggaa aaaggtcgaa aaggtctctt tcctgtggat agcacgtaca 540
ttgggaaccc aaagccgtac attgggaacc ggaacccgta catggggaac ccaaagccgt 600
acattgggaa ccgggtcacac atgtaagtga ctgatataaa agagaaaaaa ggcgattttt 660
ccgcctaata ctctttaaaa ctatttaaaa ctcttaaaac ccgcctggcc tgtgcataac 720
tgtctggcca gcgcacagcc gaagagctgc aaaaagcgcc tacccttcgg tcgctgcgct 780
ccctacgccc cgccgcttcg cgtcggccta tcgcggccgc tggccgctca aaaatggctg 840
gcctacggcc aggcaatcta ccagggcgcg gacaagccgc gccgtcgcca ctcgaccgc 900

```

ggcgccca	tcaaggcacc	ctgcctcgcg	cgtttcgggtg	atgacgggtga	aaacctctga	960
cacatgcagc	tcccgagagac	ggtcacagct	tgtctgtaag	cggatgcccgg	gagcagacaa	1020
gcccgtcagg	gcgcgtcagc	gggtgttggc	gggtgtcggg	gcgcagccat	gacccagtca	1080
cgtagcgata	gcggagtgtg	tactggctta	actatgcggc	atcagagcag	attgtactga	1140
gagtgcacca	tatgcgggtgt	gaaataccgc	acagatgcgt	aaggagaaaa	taccgcatca	1200
ggcgctcttc	cgcttcctcg	ctcactgact	cgctgcgctc	ggtcgttcgg	ctgcggcgag	1260
cggtatcagc	tcactcaaag	gcggtaatat	ggttatccac	agaatcaggg	gataacgcag	1320
gaaagaacat	gtgagcaaaa	ggccagcaaa	aggccaggaa	ccgtaaaaag	gccgcgttgc	1380
tggcggtttt	ccataggctc	gccccccctg	acgagcatca	caaaaaatcg	cgctcaagtc	1440
agaggtggcg	aaacccgcag	ggactataaa	gataccaggc	gtttccccct	ggaagctccc	1500
tcgtgcgctc	tcctgttccg	accctgcccg	ttaccggata	cctgtccgcc	tttctccctt	1560
cgggaagcgt	ggcgctttct	catagctcac	gctgtaggta	tctcagttcg	gtgtaggctcg	1620
ttcgctccaa	gctgggctgt	gtgcacgaac	cccccggttca	gcccagaccgc	tgcgcccttat	1680
ccggtaaact	tcgtcttgag	tccaacccgg	taagacacga	cttatcgcca	ctggcagcag	1740
ccactggtaa	caggattagc	agagcgaggt	atgtacggcg	tgctacagag	ttcttgaagt	1800
ggtggcctaa	ctacggctac	actagaagga	cagtatttgg	tatctgcgct	ctgctgaagc	1860
cagttacctt	cggaaaaaga	gttggtagct	cttgatccgg	caaaacaaacc	accgctggta	1920
gggtgtgttt	ttttgtttgc	aagcagcaga	ttacgcgcag	aaaaaaaagga	tctcaagaag	1980
atcctttgat	cttttctacg	gggtctgacg	ctcagtgga	cgaaaactca	cgttaaggga	2040
ttttggctat	gcattctagg	tactaaaaa	attcatccag	taaaaatata	tatttttatt	2100
tctcccaatc	aggcttgatc	cccagtaagt	caaaaaaatag	ctcgacatac	tgttcttccc	2160
cgatatccct	cctgatcgac	cggacgcaga	aggcaatgtc	ataccacttg	tccgccctgc	2220
cgtctctccc	aagatcaata	aagccactta	ctttgccatc	tttcacaaag	atgttgctgt	2280
ctcccaggtc	gccgtgggaa	aagacaagtt	cctcttcggg	cttttccgtc	tttaaaaaat	2340
catacagctc	gcgcggatct	ttaaatggag	tgtcttcttc	ccagttttcg	caatccacat	2400
gggccagatc	gttattcagt	aagtaatcca	attcggttaa	gcggtgtgtc	aagctattcg	2460
tatagggaca	atccgatatg	tcgatggagt	gaaagagcct	gatgcactcc	gcatacagct	2520
cgataatctt	ttcagggctt	tgttcatctt	catactcttc	cgagcaaaag	acgccatcgg	2580
cctcactcat	gagcagattg	ctccagccat	catgccgttc	aaagtgcagg	acctttggaa	2640
caggcagctt	tccttccagc	catagcatca	tgctcttttc	ccgttccaca	tcatagggtg	2700
tccctttata	ccggctgtcc	gtcattttta	aatataggtt	ttcattttct	cccaccagct	2760
tatataccct	agcaggagac	attccttccg	tatcttttac	gcagcggtat	ttttcgatca	2820
gttttttcaa	gttcggtgat	attctcattt	tagccattta	ttattttcct	cctctttctt	2880
acagtattta	aagatacccc	aagaagctaa	ttataacaag	acgaactcca	attcactggt	2940
ccttgcattc	taaaacctta	aataccagaa	aacagctttt	tcaaagttgt	tttcaaagtt	3000
ggcgataaac	atagatcga	cggagccgat	tttgaaccgc	cggtgatcac	aggcagcaac	3060
gctctgtcat	cgttacaatc	aacatgctac	cctccgcgag	atcatccgtg	tttcaaacc	3120
ggcagcttag	ttgccgttct	tccgaatagc	atcggttaaca	tgagcaaaag	ctgccgcctt	3180
acaacggctc	tcctgcgtac	gccgtcccgc	actgatgggc	tgctgtatc	gagtggtg	3240
tttgtgccga	gctgccggtc	ggggagctgt	tggtggctg	gtggcaggat	atattgtggt	3300
gtaaacaaat	tgacgcttag	acaacttaat	aacacattgc	ggacgttttt	aatgtactga	3360
attaacgcgc	aattaatcgc	ggggatctgg	attttagtag	tggtattttg	tttttagga	3420
tagaaaattt	attgatagaa	gtattttaca	aatacaaata	cataactaagg	gtttcttata	3480
tgctcaacac	atgagcgaaa	ccctatagga	accctaattc	ccttatctgg	gaactactca	3540
caactattta	tggagaaact	cgagtcaaat	ctcggtagcg	ggcaggaccg	gacggggcgg	3600
taccggcagg	ctgaagtcca	gctgccagaa	acccacgtca	tgccagttcc	cgtgcttgaa	3660
gcccggccgc	cgagcatgac	cgccgggggc	atatccgagc	gcctcgtagc	tgccgacgct	3720
cgggtcggtg	ggcagcccga	tgacagcgac	cacgctcttg	aagccctgtg	cctccaggga	3780
cttcagcagg	tggtgtaga	gcgtggagcc	cagtcccgtc	cgctgggtgg	ggggggagac	3840
gtacacggtc	gactcgcccg	tccagtcgta	ggcggtgctg	gccttccagg	ggcccgcgta	3900
ggcgatgccg	gcgacctcgc	cgtccacctc	ggcgacgagc	cagggatagc	gctccgcgag	3960
acggacgagg	tcgtccgtcc	actcctgcgg	ttcctgcggc	tcggtagcga	agttgaccgt	4020
gcttgtctcg	atgtagtgg	tgacgatgg	gcagaccgcc	ggcatgtccg	cctcggtggc	4080
acggcggatg	tcggccgggc	gtcgttctgg	gctcatggta	gactcgagag	agatagattt	4140
gtagagagag	actgggtgatt	tcagcgtgtc	ctctccaaat	gaaatgaact	tccttatata	4200
gaggaaggct	ttgcgaagga	tagtgggatt	gtgcgtcatc	ccttacgtca	gtggagatat	4260
cacatcaate	cacttgcttt	gaagacgtgg	ttggaacgtc	ttctttttcc	acgatgctcc	4320
tcgtgggtgg	gggtccatct	ttgggaccac	tgctggcgaga	ggcatcttga	acgatagcct	4380
ttcctttatc	gcaatgatgg	catttgtagg	tgccaccttc	cttttctact	gtccttttga	4440
tgaagtgaca	gatagctggg	caatggaatc	cgaggaggtt	tcccgatatt	accttttgtt	4500
gaaaagttct	aatagccctt	tggtcttctg	agactgtatc	tttgatattc	ttggagtaga	4560
cgagagtgtc	gtgctccacc	atgttatcac	atcaatccac	ttgctttgaa	gacgtggttg	4620
gaacgtcttc	tttttccacg	atgctcctcg	tgggtggggg	tccatctttg	ggaccactgt	4680
cggcagaggg	atcttgaacg	atagccttcc	ctttatcgca	atgatggcat	ttgtaggtgc	4740
caccttcctt	ttctactgtc	cttttgatga	agtgacagat	agctgggcaa	tggaaatccga	4800
ggaggtttcc	cgatatacc	ctttgttgaa	aagtctcaat	agccctttgg	tcttctgaga	4860
ctgtatcttt	gatattcttg	gagtagacga	gagtgctgtg	ctccaccatg	ttggcaagct	4920

gctctagcca	atacgcaaac	cgcctctccc	cgcgcgttgg	ccgattcatt	aatgcagctg	4980
gcacgacagg	tttcccgcact	ggaaaagcggg	cagtgcgcgc	aacgcaatta	atgtgagtta	5040
gctcactcat	taggcacccc	aggctttaca	ctttatgctt	ccggctcgtg	tggtgtgtgg	5100
aattgtgagc	ggataacaat	ttcacacagg	aaacagctat	gaccatgatt	acgaattcga	5160
gccttgacta	gagggtcgac	ggtatacaga	catgataaga	tacattgatg	agtttggaca	5220
aaccacaact	agaatgcagt	gaaaaaaatg	ctttatttgt	gaaatttgtg	atgctattgc	5280
tttatttgta	accattataa	gctgcaataa	acaagttggg	gtgggcgaag	aactccagca	5340
tgagatcccc	gcgctggagg	atcatccagc	cggcgtcccc	gaaaacgatt	ccgaagccca	5400
acctttcata	gaaggcggcg	gtggaatcga	aatctcgtag	cacgtgtcag	tcctgctcct	5460
cggccacgaa	gtgcacgcag	ttgccggccg	ggtcgcgcag	ggcgaactcc	cgccccacg	5520
gctgctcgcc	gatctcggtc	atggccggcc	cggaggcgctc	ccggaagtcc	gtggacacga	5580
cctccgacca	ctcggcgtag	agctcgtcca	ggcgcgcgac	ccacaccag	gccagggtgt	5640
tgctccggcac	cacctgggtcc	tggaaccgcgc	tgatgaacag	ggtcacgtcg	tcccggacca	5700
caccggcgaa	gtcgtcctcc	acgaagtccc	gggagaaccc	gagccggctcg	gtccagaaat	5760
cgaccgctcc	ggcgcagctcg	cgcgcgggtga	gcaccgggaac	ggcactgggtc	aacttggcca	5820
tggtatccaga	tttcgctcaa	gttagtataa	aaaagcaggc	ttcaatcctg	caggaattcg	5880
atcgacactc	tcgtctactc	caagaatatac	aaagatacag	tctcagaaga	ccaaagggct	5940
attgagactt	ttcaacaaag	ggtaatatcg	ggaaacctcc	tcggattcca	ttgccagct	6000
atctgtcact	tcataaaaag	gacagtagaa	aaggaagggtg	gcacctacaa	atgccatcat	6060
tgcgataaag	gaaaggctat	cggtcaagat	gcctctgccc	acagtgggtcc	caaagatgga	6120
cctccgacca	cgaggagcat	cgtggaaaaa	gaagacgttc	caaccacgtc	ttcaaggcaa	6180
gtggattgat	gtgataacat	ggtggagcac	gacactctcg	tctactccaa	gaatatcaaa	6240
gatacagctc	cagaagacca	aagggtctatt	gagacttttc	aacaaagggt	aatatcgga	6300
aacctcctcg	gattccattg	cccagctatc	tgctacttca	tcaaaaggac	agtagaaaag	6360
gaagggtggca	cctacaaatg	ccatcattgc	gataaaggaa	aggctatcgt	tcaagatgcc	6420
tctgcccaga	gtggtcccaa	agatggaccc	ccaccacga	ggagcatcgt	ggaaaaagaa	6480
cacgttccaa	ccacgtcttc	aaagcaagtg	gattgatgtg	atatctccac	tgacgtaagg	6540
gatgacgcac	aatcccacta	tccttcgcaa	gaccttctc	tatataagga	agttcatttc	6600
atttggagag	gacacgctga	aatcaccagt	ctctctctac	aaatctatct	ctctcgagct	6660
ttcgcgatc	cgggggggca	atgagatatg	aaaaagcctg	aactcacgcg	gacgtctgtc	6720
gagaagtttc	tgatcgaaaa	gttcgacagc	gtctccgacc	tgatgcagct	ctcggaggggc	6780
gaagaatctc	gtgctttcag	cttcgatgtg	ggaggggcgtg	gatatgtcct	gcgggtaaat	6840
agctgcgcgg	atgggtttcta	caaagatcgt	tatgtttatc	ggcactttgc	atcggcgcgg	6900
ctcccgattc	cggaaagtgt	tgacattggg	gagtttagcg	agagcctgac	ctattgcatc	6960
tcccgcctgt	cacagggtgt	cacgttgcaa	gacctgcctg	aaaccgaact	gcccgtgtgt	7020
ctacaaccgg	ctcgcggaggc	tatggatcgg	atcgtcgcgg	ccgatcttag	ccagcagagc	7080
gggttcgggc	cattcggacc	gcaagggaatc	gggtcaataca	ctacatggcg	tgattttcata	7140
tgcgcgattg	ctgatcccca	tggtgatcac	tggtcaaaactg	tgatggacga	caccgtcagt	7200
cgcctcgtcg	cgcaggctct	cgatgagctg	atgctttggg	ccgaggactg	ccccgaagtc	7260
ggcacctcg	tgacgcggga	tttcggctcc	aacaatgtcc	tgacggacaa	tgcccgcata	7320
acagcgggtca	ttgactggag	cgaggcgatg	ttcggggatt	cccaatacga	ggtcgccaac	7380
atcttcttct	ggaggccgtg	gttggcttgt	atggagcagc	agacgcgcta	cttcgagcgg	7440
aggtatccgg	agcttgacag	atcgccacga	ctccgggcgt	atatgctccg	cattgtgtct	7500
gaccaactct	atcagagctt	ggttgacggc	aatttcgatg	atgcagcttg	ggcgcagggt	7560
cgatgcgacg	caatcgtccg	atccggagcc	gggactgtcg	ggcgtacaca	aatcgccccg	7620
agaagcgcgg	ccgtctggac	cgatggctgt	gtagaagtac	tcgcccgatag	tggaaccgga	7680
cgcgccagca	ctcgtccgag	ggcaaagaaa	tagagttagt	gccgaccgga	tctgtogatc	7740
gacaagctcg	agttttctcca	taataatgtg	tgagttagttc	ccagataagg	gaattaggggt	7800
tcctataggg	tttcgctcat	gtgttgagca	tataagaaac	ccttagtatg	tatttgtatt	7860
tgtaaaatac	ttctatcaat	aaaattttcta	attcctaaaa	ccaaaatcca	gtactaaaaat	7920
ccagatcccc	cgaatttaatt	cggcgttaat	tcagatcaag	cttggcactg	gccgtogttt	7980
tacaacgtcg	tgactgggaa	aacctggcg	ttaccceaact	taatcgctt	gcagcacatc	8040
cccctttcgc	cagctggcgt	aatagcgaag	aggccccgac	cgatcgccct	tcccacagct	8100
tgcgcagcct	gaatggcgaa	tgctagagca	gcttgagctt	ggatcagatt	gtcgtttccc	8160
gccttcagtt	tggggatcct	ctagactgaa	ggcgggaaac	gacaatctga	tcatgagcgg	8220
agaattaagg	gagtcacgtt	atgacccccg	ccgatgacgc	gggacaagcc	gttttacgtt	8280
tggaactgac	agaaccgcaa	cgttgaagga	gccactcagc	cgcgggtttc	tggagtttaa	8340
tgagctaagc	acatacgtca	gaaaccatta	ttgcgcgttc	aaaagtgcgc	taaggctact	8400
atcagctagc	aaatatttct	tgtcaaaaat	gtccactga	cgttccataa	attccccctg	8460
gtatccaatt	agagttctcat	attcactctc	aatccaaata	atctgcaccg	gatctcgaga	8520
atcgaattcc	cgcggccgccc	atggtagatc	cttagtagta	aggagaagaa	ctttctact	8580
gagttgtccc	aattcttgtt	gaattagatg	gtgatgttaa	tgggcacaaa	ttttctgtca	8640
gtggagaggg	tgaaggtgat	gcaacatacg	gaaaacttac	ccttaaattt	atttgcacta	8700
ctggaaaact	acctgttcgg	tggccaacac	ttgtcactac	tttctcttat	gggtttcaat	8760
gcttttcaag	ataccagat	catatgaagc	ggcacgactt	cttcaagagc	gccatgcctg	8820
agggatagct	gcaggagagg	accatcttct	tcaaggacga	cgggaactac	aagacacgtg	8880
ctgaagtcaa	gtttgagggg	gacacctcgc	tcaacaggat	cgagcttaag	ggaatcgatt	8940

-56-

tcaaggaggga cggaaacatc ctcgggccaca agttggaata caactacaac tcccacaacg 9000
tatacatcat ggccgacaag caaaagaacg gcatcaaagc caacttcaag acccgccaca 9060
acatcgaaga cggcgccgtg caactcgctg atcattatca acaaaatact ccaattggcg 9120
atggccctgt cctttttacca gacaaccatt acctgtccac acaatctgcc ctttcgaaag 9180
atcccaacga aaagagagac cacatggtcc ttcttgagtt tgtaacagct gctgggatta 9240
cacatggcat ggatgaacta tacaaagcta gccaccacca ccaccaccac gtgtgaattg 9300
gtgaccagct cgaatttccc cgatcggttca aacatttggc aataaaagttt cttaaagattg 9360
aatcctgttg ccggtcttgc gatgattatc atataatttc tgttgaatta cgттааgсat 9420
gтаатаатта catgacgtta ttatgagat gggtttttat gattagagtc 9480
ccgcaattat acatttaata cgcgatagaa aacaaaatat agcgcgcaaa ctaggataaa 9540
ttatcgcgcg cgggtgtcatc tatgttacta gatcgggaat taaactatca gtgtttgaca 9600
ggatatattg cggggtaaac ctaagagaaa agagcggtta ttagaataac ggatatttaa 9660
aagggcggtg aaagggttat ccggttcgtcc atttgtatgt gcatgccaac cacagggttc 9720
ccctcgggat caaagtaact cctccgctgc tatagtgcag tcggcttctg 9780
acgttcagtg cagcgtctt ctgaaaacga catgtcgac aagtcctaag ttacgcgaca 9840
ggctgcgcgc ctgccccttt cctggcggtt tcttgtcgcg tgttttagtc gcataaagta 9900
gaataactgc gactagaacc ggagacatta cgccatgaac aagagcgccg ccgctggcct 9960
gctgggttat gcccggtca gcaccgacga ccaggacttg accaaccaac gggcggaact 10020
gcacgcggcc ggctgcacca agctgttttc cgagaagatc accggcacca ggcgcgaccg 10080
cccgagctg gccaggatgc ttgaccacct acgcccctggc gacgttgtga cagtgaccag 10140
gctagaccgc ctggcccgca gcacccgga cctactggac attgccgagc gcattccagg 10200
ggccggcgcg ggctgcgta gcctggcaga gccgtgggccc gacaccacca cgcggcgccg 10260
ccgcatgggt ttgaccgtgt tcgcccgcac tgccgagttc gagcgttccc таатсатсa 10320
ccgcacccg agcgggcgcg aggcggccaa ggcccgaggc gtgaagtttg gccccgcgcc 10380
tacctcacc ccggcacaga tcgcgcacgc ccgcgagctg atcgaccagg aaggccgcac 10440
cgtgaaagag gccgctgcac tgcttggcgt gcatcgctcg accctgtacc gcgcacttga 10500
gcgcagcgag gaagtgcgc ccaccgagcg caggcgcgcc ggtgccttcc gtgaggacgc 10560
attgaccgag gccgacgccc tggcgccgc cgagaatgaa cgccaagagg aacaagcatg 10620
aaaccgcacc aggcaggcca ggacgaaccg tttttcatta ccgaagagat cgaggcgag 10680
atgatcgcg gggttacgt gttcgagccg cccgcgcacg tctcaaccgt ctcggctgag 10740
gaaatcctgg ccggtttgtc tgatgccaa ctggcgccct ggccggccag cttggccgct 10800
gaagaaaccg agcgccgcgg tctaaaaagg tgatgtgtat ttgagtaaaa cagcttgctg 10860
gtcgcggtcg ctgctatat gatgcgatga gtaataaac aaatacgcaa ggggaacgca 10920
tgaaggltat cgtgtactt aaccagaaag cggggtcagg caagacgacc atcgcaacc 10980
atctagcccg cgccttgcga ctgcgcgggg ccgatgttct gttagtcgat tccgatcccc 11040
agggcagtg cgcgattgg gccgctgcg aggaagatca accgctaacc gttgtcggca 11100
tcgaccgccc gacgattgac cgcgacgtga aggccatcg ccggcgcgac ttctgtagta 11160
tcgacgggag gccccaggcg gcggacttgg ctgtgtccgc gatcaaggca gccgacttgc 11220
tgctgattcc ggtgcagcca agcccttacg acatattggc caccgccgac ctggtggagc 11280
tggttaagca gtcgattgag gtcacggatg gaaggctaca agcggccttt gtcgtgtcgc 11340
gggcgatcaa aggcacgcgc atcgccgggtg aggttgccga ggcgctggcc gggtagcagc 11400
tgcccattct tgagtcocgt atcacgcagc gcgtgagcta cccaggcact gcccgccgcg 11460
gcacaaccgt tcttgaatca gaaccgagg gcgacgctgc ccgcgaggtc caggcgctgg 11520
ccgctgaaat таатсaaaa ctcatttgag ttaatgaggt aaagagaaaa tgagcaaaag 11580
cacaaacacg ctaagtgcgc gccgtccgag cgcacgcagc agcaaggctg caacgttggc 11640
cagcctggca gacacgccc ccatgaagcg ggtcaacttt cagttgcccg cggaggatca 11700
caccaagctg aagatgtacg cggtagccca aggcagacc attaccgagc tgctatctga 11760
atacatcgcg cagctaccag agtaaatgag caaatgaata aatgagtaga tgaattttag 11820
cggctaaagg aggcggcatg gaaaatcaag aacaaccagg caccgacgcc gtggaatgcc 11880
ccatgtgtgg aggaacgggc ggttggccag gcgtaagcgg ctgggttgtc tgccggccct 11940
gcaatggcac tggaaaccccc aagcccaggg aatcgccgtg acggtcgcaa accatccggc 12000
ccggtacaaa tcggcgcggc gctgggtgat gacctggtgg agaagttgaa ggccgcgcag 12060
gccgcccagc ggcaacgcat cgaggcagaa gcacgccccg gtgaatcgtg gcaagcggcc 12120
gctgatcgaa tccgcaaaga atcccggcaa ccgcggcgag ccggtgcgcc gtcgattagg 12180
aagccgccc aagggcgacga gcaaccagat ttttctgttc cgatgctcta tgacgtgggc 12240
accgcgata gtcgcagcat catggacgtg gccgttttcc gtctgtcgaa gcgtgaccga 12300
cgagctggcg aggtgatccg ctacgagctt ccagacgggc acgtagaggt ttccgcagg 12360
ccggccggca tggccagtg gtgggttac gcactgggtac tgatggcggt tccccacta 12420
accgaatcca tgaaccgata ccgggaagg ccgggaagaca agcccggccg cgtgttccgt 12480
ccacacgttg cggacgtact caagttctgc cggcgagccg atggcggaag gcagaaagac 12540
gacctggtag aaacctgcat tcggttaaac accacgcacg ttgccatgca gc 12592

<210> 96

<211> 3357

<212> DNA

<213> Artificial Sequence

-57-

<220>

<223> pGEMEasyNOS Plasmid

<400> 96

tatcactagt	gaattcgcgg	ccgcctgcag	gtcgaccata	tgggagagct	cccaacgcgt	60
tggatgcata	gcttgagtat	tctatagtg	cacctaaata	gcttggcgta	atcatgggtca	120
tagctgtttc	ctgtgtgaaa	ttgttatccg	ctcacatttc	cacacaacat	acgagccgga	180
agcataaagt	gtaaagcctg	gggtgcctaa	tgagtgcgt	aactcacatt	aattgcgttg	240
cgctcactgc	ccgctttcca	gtcgggaac	ctgtcgtgcc	agctgcatta	atgaatcggc	300
caacgcgcgg	ggagaggcgg	tttgcgtatt	gggcgcctct	ccgcttcctc	gctcactgac	360
tcgctgcgct	cggctcgctc	gctgcggcga	gcggtatcag	ctcactcaaa	ggcggtaata	420
cggttatcca	cagaatcagg	ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	480
aaggccagga	accgtaaaaa	ggccgcgttg	ctggcggttt	tccataggct	ccgccccct	540
gacgagcatc	acaaaaatcg	acgctcaagt	cagaggtggc	gaaacccgac	aggactataa	600
agataccagg	cgtttccccc	tggaagctcc	ctcgtgcgct	ctcctgttcc	gaccctgccg	660
cttaccggat	acctgtccgc	ctttctccct	tcgggaagcg	tggcgctttc	tcatagctca	720
cgctgtaggt	atctcagttc	gggtgtaggtc	gttcgctcca	agctgggctg	tgtgcacgaa	780
ccccccgttc	agcccgaccg	ctgcgcctta	tcgggtaact	atcgtcttga	gtccaaacccg	840
gtaagacacg	acttatcgcc	actggcagca	gccactggta	acaggattag	cagagcgagg	900
tatgtaggcg	gtgctacaga	gttcttgaag	tggtggccta	actacggcta	cactagaaga	960
acagattttg	ctatctgcgc	tctgctgaag	ccagttacct	tcggaaaaaag	agtgtgtagc	1020
tcttgatccg	gcaaacaac	caccgctggt	agcgggtggt	tttttggttg	caagcagcag	1080
attacgcgca	gaaaaaaagg	atctcaagaa	gatcctttga	tcttttctac	ggggtctgac	1140
gctcagtggg	acgaaaactc	acgttaaggg	attttggtca	tgagattatc	aaaaagggatc	1200
ttcacctaga	tcctttttaa	ttaaaaatga	agtttttaaat	caatctaaag	tatatatgag	1260
taaacttggt	ctgacagtta	ccaatgctta	atcagtgagg	cacctatctc	agcgatctgt	1320
ctatttcggt	catccatagt	tgccctgactc	cccgctggt	agataaactac	gatacgggag	1380
ggcttaccat	ctggccccag	tgctgcaatg	ataccgcgag	acccacgctc	accggctcca	1440
gattttacag	caataaacca	gccagccgga	agggccgagc	gcagaagtgg	tcctgcaact	1500
ttatcccgct	ctatccagtc	tattaattgt	tgccgggaag	ctagagtgaag	tagttcgcca	1560
gttaatatgt	tcgcgaacgt	tggtgccatt	gctacaggca	tcgtggtgtc	acgctcgctc	1620
tttggtatgg	cttcattcag	ctccggttcc	caacgatcaa	ggcgagttac	atgatcccc	1680
atggtgtgga	aaaaaacggt	tagctccttc	ggctcctcca	tcgttgtcag	aagtaagtgt	1740
gccgcagtg	tatcactcat	ggttatggca	gcactgcata	attctcttac	tgtcatgcca	1800
tcgctaagat	gcttttctgt	gactgggtgag	tactcaacca	agtcattctg	agaatagtgt	1860
atgcggcgac	cgagttgctc	ttgcccgcg	tcaataccgg	ataataccgc	gccacatgag	1920
agaactttaa	aagtgtcat	cattggaaaa	cgttcttcgg	ggcgaaaact	ctcaaggatc	1980
ttaccgctgt	tgagatccag	ttcgatgtaa	cccactcgtg	cacccaactg	atcttcagca	2040
tcttttactt	taccagcgt	ttctgggtga	gcaaaaacag	gaaggcaaaa	tgccgcaaaa	2100
aagggaataa	ggcgacacg	gaaatgttga	atactcatac	tcttcctttt	tcaatattat	2160
tgaagcattt	atcagggtta	ttgtctcatg	agcggataca	tatttgaaatg	tatttagaaa	2220
aataaacaac	taggggttcc	gcgcacattt	ccccgaaaag	tgccacctga	tcggtgtgta	2280
aataccgcac	agatgcgtaa	ggagaaaata	ccgcacatagg	aaattgtgaag	cgtaaatatt	2340
ttgttaaaat	tcgcgttaaa	tttttggtta	atcagctcat	tttttaacca	ataggccgaa	2400
atcggcaaaa	tcctttataa	atcaaaagaa	tagaccgaga	taggggttag	tggtgttcca	2460
gtttggaaca	agagtcact	attaaagaac	gtggactcca	acgtcaaagg	gcgaaaaacc	2520
gtctatcagg	gcgatggccc	actacgtgaa	ccatcaccct	aatcaagttt	tttggggctg	2580
aggtgccgta	aagcactaaa	tcggaaccct	aaaggagacc	cccgatthag	agcttgacgg	2640
ggaaagccgg	cgaacgtggc	gagaaaggaa	gggaagaaag	cgaaaggagc	ggcgctagg	2700
gcgctggcaa	gtgtagcggg	cacgctgcgc	gtaaccacca	caccgcgccg	gcttaatgcg	2760
ccgctacagg	gcgcgtccat	tcgccattca	ggctgcgcaa	ctgttgggaa	ggcgatcgg	2820
tcggggcctc	ttcgctatta	cgccagctgg	cgaaaggggg	atgtgctgca	aggcgattaa	2880
gttgggtaac	gccagggttt	tcccagtcac	gacgttgtaa	aacgacggcc	agtgaattgt	2940
aatacgaact	actatagggc	gaattggggc	cgacgtcgca	tgctcccggc	cgccatggcg	3000
gccgcgggaa	ttcgattctc	gagatccggg	gcagattatt	tggattgaga	gtgaatatga	3060
gactctaatt	ggataccgag	gggaatttat	ggaacgtcag	tggagcattt	ttgacaagaa	3120
atatgtgcta	gctgatagtg	accttagggc	acttttgaa	gcgcaataat	ggtttctgac	3180
gtatgtgctt	agctcattaa	actccagaaa	cccgcggctg	agtggctcct	tcaacgttgc	3240
ggttctgtca	gttccaaacg	taaaacggct	tgtcccgcgt	catcgccggg	ggtcataaac	3300
tgactccctt	aattctccgc	tcatgatcag	attgtcggtt	cccgccttca	gtctaga	3357

<210> 97

<211> 10122

<212> DNA

<213> Artificial Sequence

<220>

-58-

<223> p1302NOS Plasmid

<400> 97

catggttagat	ctgactagta	aaggagaaga	actttttcact	ggagttgtcc	caattcttgt	60
tgaattagat	ggtgatgta	atgggcacaa	attttctgtc	agtggagagg	gtgaagggtga	120
tgcaacatac	ggaaaactta	cccttaaatt	tatttgact	actggaaaac	tacctgttcc	180
gtggccaaca	cttgtcacta	ctttctctta	tgggtgtcaa	tgcttttcaa	gatacccgaga	240
tcatatgaag	cgggcagact	tcttcaagag	cgccatgcct	gagggatacg	tgcaggagag	300
gaccatcttc	ttcaaggacg	acgggaacta	caagacacgt	gctgaagtca	agtttgaggg	360
agacaccctc	gtcaacagga	tcgagcttaa	gggaatcgat	ttcaaggagg	acggaaacat	420
cctcgccac	aagttggaat	acaactacaa	ctcccacaac	gtatacatca	tggccgacaa	480
gcaaaagaac	ggcatcaaag	ccaacttcaa	gaccgcgcac	aacatcgaag	acggcgcgct	540
gcaactcgct	gatcattatc	aacaaaatac	tccaattggc	gatggccctg	tccttttacc	600
agacaacat	tacctgtcca	cacaatctgc	cctttcgaaa	gatcccaacg	aaaagagaga	660
ccatgatttc	cttcttgagt	ttgtaacagc	tgctgggatt	acacatggca	tggatgaact	720
atacaaagct	agccaccacc	accaccacca	cgtgtgaatt	ggtgaccagc	tcgaatttcc	780
ccgatcggtc	aaacatttgg	caataaagtt	tcttaagatt	gaatcctgtt	gccggtcttg	840
cgatgatttc	catataattt	ctgttgaatt	acgttaagca	tgtaataatt	aacattgagg	900
gcatgacggt	atattatgaga	tgggttttta	tgattagagt	cccgaatta	tacatttaat	960
acggcataga	aaacaaaata	tagcgcgcaa	actaggataa	attatcgcg	gcggtgtcat	1020
ctatgttact	agatcgggaa	ttaaactatc	agtgttgac	aggatatatt	ggcggtgtaa	1080
cctaagagaa	aagagcggtt	attagaataa	cggatattta	aaagggcggt	aaaaggttta	1140
tccgttcgtc	catttgtatg	tgcattgcaa	ccacagggtt	cccctcggg	tcaaagtact	1200
ttgatccaac	ccctcgctg	ctatagtcca	gtcggcttct	gacgttcagt	gcagccgtct	1260
tctgaaaacg	acatgtcgca	caagtccata	gttacgcgac	aggctgccc	cctgcccctt	1320
tccctggcgt	ttcttgtcgc	gtgttttagt	cgcataaagt	agaataactg	cgactagaac	1380
cggagacatt	acgccatgaa	caagagcgcc	gcgctggccc	tgctgggcta	tgcccgcgct	1440
agcaccgacg	accaggactt	gaccaaccaa	cgggcccgaac	tgacgcgggc	cggctgcacc	1500
aatgtgtttt	ccgagaagat	caccggcacc	aggcgcgacc	gcccggagct	ggccaggatg	1560
cttgaccacc	tacgcccctg	cgacttctg	acagttagca	ggctagaccg	ctggcccgct	1620
agcaccgcg	acctactgga	cattgcccag	cgcattccagg	aggccggcgc	gggctgctgt	1680
agcctggcag	agccgtgggc	cgacaccacc	acgcggcgcc	gcccgcaggt	gttgaccgtg	1740
ttcgccggca	ttgcccagtt	cgagcgttcc	ctaatctcg	accgcaccgc	gagcgggcgc	1800
gaggccgcca	aggcccagag	cgtgaagttt	ggcccccgcc	ctaccctcac	cccggcacag	1860
atcgccgacg	cccgcgagct	gatcgaccag	gaaggccgca	ccgtgaaaga	ggcggttgca	1920
ctgcttggcg	ctcatcgctc	gacctgttac	cgcgacttg	agcgagcgga	ggaagtgcag	1980
cccaccgcg	ccaggccggc	cgggtgccttc	cgtgaggacg	cattgaccga	ggccgagccc	2040
ctggccggcg	ccgagaatga	acgccaagag	gaacaagcat	gaaaccgcac	caggacggcc	2100
aggacgaacc	gtttttcatt	accgaagaga	tcgaggcgga	gatgatcgcg	gcccgggtacg	2160
tggtcgagcc	gcccgcgcac	gtctcaaccg	tgcggtgca	tgaaatcctg	gcccgtttgt	2220
ctgatgccaa	gctggcggcc	tggccggcca	gcttggccgc	tgaagaaacc	gagcgccgcc	2280
gtctaaaaag	gtgatgtgta	tttgagttaa	acagcttgcg	tcattgcggtc	gctgcgtata	2340
tgatgcgatg	agtaaaataa	caaatacgca	aggggaacgc	atgaagggtta	tcgctgtact	2400
taaccagaaa	ggcgggtcag	gcaagacgac	catcgcaacc	catctagccc	gccccttgca	2460
actcgccggg	gcccgatgtt	tgttagtcca	ttccgatccc	cagggcagtg	cccgcgattg	2520
ggcggccgtg	cggggaagatc	aaccgctaac	cgttgtcgcc	atcgaccgcc	cgacgattga	2580
ccgcgacgtg	aaggccatcg	gcccggcgca	cttcgtagtg	atcgacggag	cgcccaggcc	2640
ggcggacttg	gctgtgtccg	cgatcaaggc	agccgacttc	gtgctgattc	cgggtgcagcc	2700
tatcacgcag	cgcgtgagct	ccaccgcgca	cctggtggag	ctggttaagc	agcgattga	2760
ggtcacggat	ggaaggctac	aagcggccct	tgtcgtgtcg	cgggcgatca	aaggcacgcg	2820
catcgccggt	gaggttgccg	aggcgctggc	cgggtacgag	ctgcccattc	ttgagtcctg	2880
tatcacgcag	cgcgtgagct	acccaggcac	tgccgcggcc	ggcacaaccg	ttcttgaatc	2940
agaaccgcag	ggcgacgctg	cccgcgaggt	ccaggcgctg	gcccgtgaaa	ttaaatcaaa	3000
actcatttga	gttaattgag	taaagagaaa	atgagcaaaa	gcacaaacac	gctaagtgc	3060
ggcgtccga	gcgcacgcag	cagcaaggct	gcaacgttgg	ccagcctggc	agacacgcca	3120
gccatgaagc	gggtcaactt	tcagttgccg	gcgaggatc	acaccaagct	gaagatgtac	3180
gcggtacgcc	aaggcaagac	cattaccgag	ctgctatctg	aatacatcgc	gcagctacca	3240
gagtaaatga	gcaaatgaat	aaatgagtga	atgaatttta	gcggtctaaag	gagcgccgat	3300
ggaaaatcaa	gaacaaccag	gcaccgacgc	cgtggaatgc	cccatgtgtg	gaggaacggg	3360
cggttggcca	ggcgtaagcg	gctgggttgt	ctgcgggccc	tgcaatggca	ctggaacccc	3420
caagtcggag	gaatcggcgt	gacggtcgca	aaccatccgg	cccggtaaaa	atcgccgctg	3480
cgctgggtga	tgacctgggt	gagaagttga	aggccgcgca	ggccgcccag	cggcaacgca	3540
tcgaggcaga	agcacgcccc	ggtgaatcgt	ggcaagcggc	cgctgatcga	atccgcaaa	3600
aatcccgcca	acccgcggca	gcccgtgcgc	cgctgattag	gaagccgccc	aagggcgagc	3660
agcaaccaga	ttttttcgtt	ccgatgctct	atgacgtggg	caccgcgcat	agtcgcagca	3720
tcatggacgt	ggccggtttc	cgtctgtcga	agcgtgaccg	acgagctggc	gaggtgatcc	3780
gctacgagct	tccagacggg	cacgtagagg	tttccgcagg	gcccgcgggc	atggccagtg	3840

tgtgggatta	cgacctggta	ctgatggcgg	tttcccatct	aaccgaatcc	atgaaccgat	3900
accgggaagg	gaagggagac	aagcccgcc	gcgtgttccg	tccacacgtt	gcggacgtac	3960
tcaagttctg	ccggcgagcc	gatggcggaa	agcagaaaga	cgacctggta	gaaacctgca	4020
ttcgggttaa	caccacgcac	gttgccatgc	agcgtacgaa	gaaggccaag	aacggccgcc	4080
tggtagcggg	atccgagggg	gaagccttga	ttagccgcta	caagatcgta	aagagcgaaa	4140
ccgggcggcc	ggagtacatc	gagatcgagc	tagctgattg	gatgtaccgc	gagatcacag	4200
aaggcaagaa	cccggacgtg	ctgacgggtc	accccgatta	ctttttgatc	gatcccggca	4260
tcggccggtt	tctctaccgc	ctggcacgcc	gcgcgcgagg	caaggcagaa	gccagatggg	4320
tgttcaagac	gatctacgaa	cgcagtggca	gcgcgcggaga	gttcaagaag	ttctgtttca	4380
ccgtgcgcaa	gctgatcggg	tcaaatgacc	tgccggagta	cgatttgaag	gaggaggcgg	4440
ggcaggctgg	cccgatcccta	gtcatgcgct	accgcaacct	gategagggc	gaagcatccg	4500
ccggttccta	atgtacggag	cagatgctag	ggcaaatgtc	cctagcaggg	gaaaaagggtc	4560
gaaaagggtc	ctttcctgtg	gatagcacgt	acattgggaa	cccaaagccg	tacattggga	4620
accggaaccc	gtacattggg	aacccaaagc	cgtacattgg	gaaccgggtc	cacatgtaag	4680
tgactgatat	aaaagagaaa	aaaggcgatt	tttccgccta	aaactcttta	aaactcttta	4740
aaactcttaa	aaccgcgctg	gcctgtgcat	aactgtctgg	ccagcgcaca	gccgaagagc	4800
tgcaaaaagg	gcctaccctt	cggtcgctgc	gctccctacg	ccccgcgctc	tcgctgcggc	4860
ctatcgccgg	cgctggccgc	tcaaaaatgg	ctcgccctacg	gccaggcaat	ctaccagggc	4920
gcggacaagc	cgcgccgctc	ccactcgacc	gccggcgccc	acatcaaggc	accctgcctc	4980
gcgcggttct	gtgatgacgg	tgaaaacctc	tgacacatgc	agctcccggg	gacgggtcaca	5040
gcttgctgtg	aagcggatgc	cgggagcaga	caagcccgctc	agggcgcgctc	agcgggtggt	5100
ggcgggtgtc	ggggcgcgagc	catgacccag	tcacgtagcg	atagcggagt	gtatactggc	5160
ttactatgc	ggcatcagag	cagattgtac	tgagagtgc	ccatattcg	tgtgaaatac	5220
cgcacagatg	cgtaaggaga	aaataccgca	tcaggcgctc	ttccgcttcc	tcgctcactg	5280
actcgctgog	ctcggtcggt	cggtcgcgcc	gagcgggtatc	agctcactca	aaggcggtaa	5340
tacggttatc	cacagaatca	ggggataacg	caggaagaa	catgtgagca	aaaggccagc	5400
aaaaggccag	gaaccgtaaa	aaggccgctg	tttccatagg	tttccatagg	ctccgcccc	5460
ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagagggtg	gcgaaacccg	acaggactat	5520
aaagatacca	ggcggttccc	cctggaagct	ccctcgctgc	ctctcctggt	ccgacccctgc	5580
cgcttaccgg	atacctgtcc	gccttctccc	cttcgggaag	cggtggcgctt	tctcatagct	5640
cacgctgtag	gtatctcagt	tcggtgttagg	tcgttcgctc	caagctgggc	tgtgtgcacg	5700
aaccccccg	tcagcccgac	cgctgcgcct	catccggtaa	ctatcgtctt	gagtccaacc	5760
cggttaagca	cgacttatcg	ccactggcag	tacgccattg	taacaggatt	agcagagcga	5820
ggtagtgtag	cggtgctaca	gagttcctga	agtgggtggc	taactacggc	tacacagaaa	5880
ggacagatatt	tggtatctgc	gctctgtcta	agccagttac	cttcggaaaa	agagttggta	5940
gctcttgatc	cggaacaa	accaccgctg	gtagcgggtg	tttttttggt	tgcaagcagc	6000
agattacgcg	cagaaaaaaa	ggatctcaag	aagatccttt	gatcttttct	acgggggtctg	6060
acgctcagtg	gaacgaaaaa	tcacgttaag	ggatttttgg	catgcattct	aggtactaaa	6120
acaattcatc	cagtaaaaata	taatttttta	tttttccca	atcaggcttg	atccccagtc	6180
agtcaaaaaa	tagctcgaca	tactgttctt	ccccgatatc	ctccctgatc	gaccggagcg	6240
agaaggcaat	gtcataccac	ttgtccgccc	tgccgcttct	cccaagatca	ataaagccac	6300
ttactttgccc	atcttttcaca	aagatgttgc	tgtctcccag	gtcgcgctgg	gaaaagacaa	6360
gttctctctc	gggcttttcc	gtctttaaaa	aatcatacag	cttcgcgcgga	tcttttaaat	6420
gagtgtcttc	ttcccagttt	tcgcaatcca	catcggccag	atcgttatct	agtaagtaat	6480
ccaattcggc	taagcggctg	tctaagctat	tcgtataggg	acaatccgat	atgtcgatgg	6540
agtgaagag	cctgatgcac	tcgcgataca	gctcgataat	cttttcaggg	ctttgttcat	6600
cttcatactc	ttccgagcaa	aggacgccat	cggcctcact	catgagcaga	ttgtctccagc	6660
catcatgccg	ttcaaagtgc	aggacctttg	gaacaggcag	ctttccttcc	agccatagca	6720
tcagtgcctt	ttcccgttcc	acatcatagg	tggtcccttt	ataccggctg	tcggtcattt	6780
ttaaatatag	gttttcattt	tctcccacca	gcttatatac	cttagcagga	gacattcctt	6840
ccgtatcttt	tacgcagcgg	tatttttctga	tcagtttttt	caattccggg	gatatttctca	6900
ttttagccat	ttattatttc	cttccctctt	tctacagtat	ttaaagatac	cccaagaagc	6960
taattataac	aagacgaact	ccaattcact	gttccctgca	ttctaaaacc	ttaaatacca	7020
gaaaacagct	ttttcaaagt	tgttttcaaa	gttggcgtat	aacatagtat	cgacggagcc	7080
gatttttga	ccgcgggtgat	cacaggcagc	aacgctctgt	catcgttaca	atcaacatgc	7140
taccctccgc	gagatcatcc	gtgtttcaaa	cccggcagct	tagttgccgt	tcttccgaat	7200
agcatcggtg	acatgagcaa	agtctgcgcg	cttataacgg	ctctcccgtc	gacgcgctcc	7260
cgactgatg	ggctgcctgt	atcgagtggg	gattttgtgc	cgagctgccc	gtcggggagc	7320
tggttggtgg	ctggtggcag	gatataattgt	gggtgtaaca	aattgacgct	tagacaactt	7380
aataacacat	tgcggacggt	tttaattgtac	tgaattaacg	ccgaattaat	tcggggggatc	7440
tggttttag	tactggattt	tggttttagg	tttattgata	aaagtatttt	gaagtatttt	7500
acaaatacaa	atacatacta	agggtttctt	atatgctcaa	cacatgagcg	aaaccctata	7560
ggaaccctaa	ttcccttatc	tgggaaactac	tcacacatta	ttatggagaa	actcgagctt	7620
gtcgatcgac	agatccggtc	ggcatctact	ctatttcttt	gccctcggac	gagtgcggg	7680
gcgtcggttt	ccactatcgg	cgagtacttc	tacacagcca	tcggtccaga	cggccgcgct	7740
tctgcgggcg	atthgtgtac	gcccagacgt	cccggctccg	gatcggacga	ttgcgtcgca	7800
tcgacctgc	gcccagctg	catcatcgaa	attgcccgtc	accaagctct	gatagagttg	7860

-60-

```

gtcaagacca atgctggagca tatacgcccc gagtcgtggc gatcctgcaa gctccggatg 7920
cctccgctcg aagtagcgcg tctgctgctc catacaagcc aaccacggcc tccagaagaa 7980
gatgtgtggcg acctcgctatt gggaatcccc gaacatcgcc tcgctccagt caatgaccgc 8040
tggtatgcgg ccattgtccg tcaggacatt gttggagccg aaatccgcgt gcacgaggtg 8100
ccggacttcg gggcagtcct cggcccaaag catcagctca tcgagagcct gcgcgacgga 8160
cgactgacg gtgtcgtcca tcacagtttg ccagtgtatc acatggggat cagcaatcgc 8220
gcatatgaaa tcacgccatg tagtgatttg accgattcct tgcggtccga atgggcccga 8280
ccgctcgtc tggctaagat cggccgcagc gatcgcatcc atagcctccg cgaccggttg 8340
tagaacagcg ggcagttcgg ttccaggcag gtcttgcaac gtgacaccct gtgcacggcg 8400
ggagatgcaa taggtcaggc tctcgctaaa ctccccaatg tcaagcactt ccggaatcgg 8460
gagcgcggcc gatgcaaagt gccgataaac ataacgatct ttgtagaaac catcggcgca 8520
gctattttacc cgcaggacat atccacgccc tcctacatcg aagctgaaag caccagattc 8580
ttcgccctcc gagagctgca tcaggctcga cgcgctgtcg aacttttcga tcagaaactt 8640
ctcgacagag gtgcggtgga gttcaggctt ttctcatctt cattgcccc ccggtatctg 8700
gaaagctcga gagataga tttgtagaga gagactggtg atttcagcgt gtcctctcca 8760
aatgaaatga acttccttat atagaggaag gtcttgcgaa ggatagtggg attgtgcgtc 8820
atcccttacg tcagtggaga tatcacatca atccacttgc tttgaagacg tgggtggaac 8880
gtctctcttt tccacgatgc tctcgtggg tgggggtcca tctttgggac cactgtcggc 8940
agaggcatct tgaacgatag cctttccttt atcgcaatga tggcatttgt aggtgccacc 9000
ttcctttctt actgtccttt tgatgaagtg acagatagct gggcaatgga atccgaggag 9060
gtttcccgat attacccttt gttgaaaagt ctcaatagcc ctttgggtctt ctgagactgt 9120
atctttgata ttcttggaat agacgagagt gtcgtgctcc accatgttat cacatcaatc 9180
cacttgcttt gaagacgtgg ttggaacgtc tcttttttcc acgatgtctc tcgtgggtgg 9240
gggtccatct ttgggaccac tgctggcaga ggcactttga acgatagcct ttcctttatc 9300
gcaatgatgg catttgtagg tgccaccttc cttttctact gtccttttga tgaagtgaca 9360
gatagctggg caatggaatc cgaggaggtt tcccgatatt accctttgtt gaaaagtctc 9420
aatagccctt tggctctctg agactgtatc tttgatattc ttggagtga cgagagtgtc 9480
gtgctccacc atgttggaat gctgctctag ccaatacgca aaccgcctct ccccgcgctg 9540
tggccgattc attaatgcag ctggcacgac aggtttcccg actggaagac gggcagtgag 9600
cgcaacgcaa ttaattgtgag ttagctcact cattagcac cccaggcttt acactttatg 9660
cttcgggctc gtatgttgtg tggaatttgt agcgataac aatttcacac aggaaacagc 9720
tatgaccatg attacgaatt cgagctcggt acccggggat cctctagact gaaggcgga 9780
aacgaacatc tgatcatgag cggagaatta agggagtcac gttatgacct cgcggcatga 9840
cgcgggacaa gccgttttac gtttggaact gacagaaccg caacgttgaa ggagccactc 9900
agcccggggt ttctggagtt taatgagcta agcacatagc tcagaaacca ttattgcgcg 9960
ttcaaaagtc gcctaaggte actatcagtt tcttgtaaaa aatgctccac ctcaatccaa 10020
tgacgttcca taaattcccc tcggtatcca attagagtct catattcact ctcaatccaa 10080
ataatctgca ccggatctcg agaatcgaat tccgcgggcc gc 10122

```

<210> 98
 <211> 621
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> N. tabacum rDNA intergenic spacer (IGS) sequence

<300>
 <308> Genbank #Y08422
 <309> 1997-10-31

```

<400> 98
gtgctagcca atgtttaaca agatgtcaag cacaatgaat gttggtggtt ggtggtcgtg 60
gctggcggtg gtggaataat gcggtggttc gagcgttagt gatcggcgat ggttgggtgt 120
tgcagcggtg tttgatatac gaatcactta ggtggttgtt cacaatggag gtgcgtcatg 180
gttatgggtg gttggtcatc tatatatatt tataataata ttaagtattt tacctatttt 240
ttacataatt ttatttaaat ttatgcattg tttgtatttt taaatagtgt ttatcgtact 300
tgttttataa aatattttat tattttatgt gtttatattt tacttgatgt attggaaatt 360
ttctccattg ttttttctat attttataata attttcttat ttttttttgt tttattatgt 420
attttttcgt tttataataa atattttatta aaaaaatat tatttttgta aaatatatca 480
tttacaatgt ttaaaagtca tttgtgaata tattagctaa gttgtacttc tttttgtgca 540
tttggtgttg tacatgtcta ttatgattct ctggccaaaa catgtctact cctgtcactt 600
gggttttttt ttttaagaca t 621

```

<210> 99
 <211> 25
 <212> DNA

-61-

<213> Artificial Sequence

<220>

<223> NTIGS-F1 Primer

<400> 99

gtgctagcca atgtttaaca agatg

25

<210> 100

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> NTIGS-R1 Primer

<400> 100

atgtcttaaa aaaaaaaacc caagtgc

28

<210> 101

<211> 233

<212> DNA

<213> Mus Musculus

<300>

<308> Genbank #V00846

<309> 1989-07-06

<400> 101

gacctggaat	atggcgagaa	aactgaaaat	cacggaaaat	gagaaataca	cacttttagga	60
cgtgaaatat	ggcgaggaaa	actgaaaaag	gtggaaaatt	tagaaatgtc	cactgttagga	120
cgtggaatat	ggcaagaaaa	ctgaaaatca	tggaaaatga	gaaacatcca	cttgacgact	180
tgaaaaatga	cgaaatcact	aaaaaacgtg	aaaaatgaga	aatgcacact	gaa	233

<210> 102

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> MSAT-F1 Primer

<400> 102

aataccgcgg aagcttgacc tggaatatcg c

31

<210> 103

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> MSAT-Ri Primer

<400> 103

ataaccgcgg agtccttcag tgtgcat

27

<210> 104

<211> 277

<212> DNA

<213> Artificial Sequence

<220>

<223> Nopaline Synthase Promoter Sequence

<300>

<308> Genbank #U09365

<309> 1997-10-17

-62-

<400> 104
 gagctcgaat ttccccgata gttcaaacaat ttggcaataa agtttcttaa gattgaatcc 60
 tgttgccggt cttgcgatga ttatcatata atttctgttg aattacgtta agcatgtaat 120
 aattaacatg taatgcatga cgttatttat gagatgggtt tttatgatta gagtcccgca 180
 attatacatt taatagcgga tagaaaacaa aatatagcgc gcaaactagg ataaattatc 240
 gcgcgcggtg tcattctatgt tactagatcg ggaattc 277

<210> 105
 <211> 1812
 <212> DNA
 <213> Escherichia coli

<220>
 <221> CDS
 <222> (1)...(1812)
 <223> Beta-Glucuronidase

<300>
 <308> Genbank #S69414
 <309> 1994-09-23

<400> 105
 atg tta cgt cct gta gaa acc cca acc cgt gaa atc aaa aaa ctc gac 48
 Met Leu Arg Pro Val Glu Thr Pro Thr Arg Glu Ile Lys Lys Leu Asp
 1 5 10 15
 ggc ctg tgg gca ttc agt ctg gat cgc gaa aac tgt gga att gat cag 96
 Gly Leu Trp Ala Phe Ser Leu Asp Arg Glu Asn Cys Gly Ile Asp Gln
 20 25 30
 cgt tgg tgg gaa agc gcg tta caa gaa agc cgg gca att gct gtg cca 144
 Arg Trp Trp Glu Ser Ala Leu Gln Glu Ser Arg Ala Ile Ala Val Pro
 35 40 45
 ggc agt ttt aac gat cag ttc gcc gat gca gat att cgt aat tat gcg 192
 Gly Ser Phe Asn Asp Gln Phe Ala Asp Ala Asp Ile Arg Asn Tyr Ala
 50 55 60
 ggc aac gtc tgg tat cag cgc gaa gtc ttt ata ccg aaa ggt tgg gca 240
 Gly Asn Val Trp Tyr Gln Arg Glu Val Phe Ile Pro Lys Gly Trp Ala
 65 70 75 80
 ggc cag cgt atc gtg ctg cgt ttc gat gcg gtc act cat tac ggc aaa 288
 Gly Gln Arg Ile Val Leu Arg Phe Asp Ala Val Thr His Tyr Gly Lys
 85 90 95
 gtg tgg gtc aat aat cag gaa gtg atg gag cat cag ggc ggc tat acg 336
 Val Trp Val Asn Asn Gln Glu Val Met Glu His Gln Gly Gly Tyr Thr
 100 105 110
 cca ttt gaa gcc gat gtc acg ccg tat gtt att gcc ggg aaa agt gta 384
 Pro Phe Glu Ala Asp Val Thr Pro Tyr Val Ile Ala Gly Lys Ser Val
 115 120 125
 cgt atc acc gtt tgt gtg aac aac gaa ctg aac tgg cag act atc ccg 432
 Arg Ile Thr Val Cys Val Asn Asn Glu Leu Asn Trp Gln Thr Ile Pro
 130 135 140
 ccg gga atg gtg att acc gac gaa aac ggc aag aaa aag cag tct tac 480
 Pro Gly Met Val Ile Thr Asp Glu Asn Gly Lys Lys Lys Gln Ser Tyr
 145 150 155 160
 ttc cat gat ttc ttt aac tat gcc gga atc cat cgc agc gta atg ctc 528
 Phe His Asp Phe Phe Asn Tyr Ala Gly Ile His Arg Ser Val Met Leu
 165 170 175
 tac acc acg ccg aac acc tgg gtg gac gat atc acc gtg gtg acg cat 576

-63-

Tyr	Thr	Thr	Pro	Asn	Thr	Trp	Val	Asp	Asp	Ile	Thr	Val	Val	Thr	His	
			180					185					190			
gtc	gcg	caa	gac	tgt	aac	cac	gcg	tct	gtt	gac	tgg	cag	gtg	gtg	gcc	624
Val	Ala	Gln	Asp	Cys	Asn	His	Ala	Ser	Val	Asp	Trp	Gln	Val	Val	Ala	
		195					200					205				
aat	ggg	gat	gtc	agc	gtt	gaa	ctg	cgt	gat	gcg	gat	caa	cag	gtg	gtt	672
Asn	Gly	Asp	Val	Ser	Val	Glu	Leu	Arg	Asp	Ala	Asp	Gln	Gln	Val	Val	
	210					215					220					
gca	act	gga	caa	ggc	act	agc	ggg	act	ttg	caa	gtg	gtg	aat	ccg	cac	720
Ala	Thr	Gly	Gln	Gly	Thr	Ser	Gly	Thr	Leu	Gln	Val	Val	Asn	Pro	His	
225					230					235					240	
ctc	tgg	caa	ccg	ggg	gaa	ggg	tat	ctc	tat	gaa	ctg	tgc	gtc	aca	gcc	768
Leu	Trp	Gln	Pro	Gly	Glu	Gly	Tyr	Leu	Tyr	Glu	Leu	Cys	Val	Thr	Ala	
				245					250					255		
aaa	agc	cag	aca	gag	tgt	gat	atc	tac	ccg	ctt	cgc	gtc	ggc	atc	cgg	816
Lys	Ser	Gln		Glu	Cys	Asp	Ile	Tyr	Pro	Leu	Arg	Val	Gly	Ile	Arg	
			260					265					270			
tca	gtg	gca	gtg	aag	ggc	gaa	cag	ttc	ctg	att	aac	cac	aaa	ccg	ttc	864
Ser	Val	Ala	Val	Lys	Gly	Glu	Gln	Phe	Leu	Ile	Asn	His	Lys	Pro	Phe	
		275					280					285				
tac	ttt	act	ggc	ttt	ggg	cgt	cat	gaa	gat	gcg	gac	ttg	cgt	ggc	aaa	912
Tyr	Phe	Thr	Gly	Phe	Gly	Arg	His	Glu	Asp	Ala	Asp	Leu	Arg	Gly	Lys	
	290					295					300					
gga	ttc	gat	aac	gtg	ctg	atg	gtg	cac	gac	cac	gca	tta	atg	gac	tgg	960
Gly	Phe	Asp	Asn	Val	Met	Val	His	Asp	His	Ala	Leu	Met	Asp	Trp		
305					310				315					320		
att	ggg	gcc	aac	tcc	tac	cgt	acc	tgc	cat	tac	cct	tac	gct	gaa	gag	1008
Ile	Gly	Ala	Asn	Ser	Tyr	Arg	Thr	Ser	His	Tyr	Pro	Tyr	Ala	Glu	Glu	
				325					330					335		
atg	ctc	gac	tgg	gca	gat	gaa	cat	ggc	atc	gtg	gtg	att	gat	gaa	act	1056
Met	Leu	Asp	Trp	Ala	Asp	Glu	His	Gly	Ile	Val	Val	Ile	Asp	Glu	Thr	
			340					345					350			
gct	gct	gtc	ggc	ttt	aac	ctc	tct	tta	ggc	att	ggg	ttc	gaa	gcg	ggc	1104
Ala	Ala	Val	Gly	Phe	Asn	Leu	Ser	Leu	Gly	Ile	Gly	Phe	Glu	Ala	Gly	
		355					360					365				
aac	aag	ccg	aaa	gaa	ctg	tac	agc	gaa	gag	gca	gtc	aac	ggg	gaa	act	1152
Asn	Lys	Pro	Lys	Glu	Leu	Tyr	Ser	Glu	Glu	Ala	Val	Asn	Gly	Glu	Thr	
		370				375					380					
cag	caa	gcg	cac	tta	cag	gcg	att	aaa	gag	ctg	ata	gcg	cgt	gac	aaa	1200
Gln	Gln	Ala	His	Leu	Gln	Ala	Ile	Lys	Glu	Leu	Ile	Ala	Arg	Asp	Lys	
					390					395					400	
aac	cac	cca	agc	gtg	gtg	atg	tgg	agt	att	gcc	aac	gaa	ccg	gat	acc	1248
Asn	His	Pro	Ser	Val	Val	Met	Trp	Ser	Ile	Ala	Asn	Glu	Pro	Asp	Thr	
				405					410					415		
cgt	ccg	caa	ggg	gca	cgg	gaa	tat	ttc	gcg	cca	ctg	gcg	gaa	gca	acg	1296
Arg	Pro	Gln	Gly	Ala	Arg	Glu	Tyr	Phe	Ala	Pro	Leu	Ala	Glu	Ala	Thr	
			420					425					430			
cgt	aaa	ctc	gac	ccg	acg	cgt	ccg	atc	acc	tgc	gtc	aat	gta	atg	ttc	1344
Arg	Lys	Leu	Asp	Pro	Thr	Arg	Pro	Ile	Thr	Cys	Val	Asn	Val	Met	Phe	
		435					440					445				

-64-

tgc	gac	gct	cac	acc	gat	acc	atc	agc	gat	ctc	ttt	gat	gtg	ctg	tgc	1392
Cys	Asp	Ala	His	Thr	Asp	Thr	Ile	Ser	Asp	Leu	Phe	Asp	Val	Leu	Cys	
	450					455					460					
ctg	aac	cgt	tat	tac	gga	tgg	tat	gtc	caa	agc	ggc	gat	ttg	gaa	acg	1440
Leu	Asn	Arg	Tyr	Tyr	Gly	Trp	Tyr	Val	Gln	Ser	Gly	Asp	Leu	Glu	Thr	
465					470					475					480	
gca	gag	aag	gta	ctg	gaa	aaa	gaa	ctt	ctg	gcc	tgg	cag	gag	aaa	ctg	1488
Ala	Glu	Lys	Val	Leu	Glu	Lys	Glu	Leu	Leu	Ala	Trp	Gln	Glu	Lys	Leu	
				485					490					495		
cat	cag	ccg	att	atc	atc	acc	gaa	tac	ggc	gtg	gat	acg	tta	gcc	ggg	1536
His	Gln	Pro	Ile	Ile	Ile	Thr	Glu	Tyr	Gly	Val	Asp	Thr		Ala	Gly	
			500					505					510			
ctg	cac	tca	atg	tac	acc	gac	atg	tgg	agt	gaa	gag	tat	cag	tgt	gca	1584
Leu	His	Ser	Met	Tyr	Thr	Asp	Met	Trp	Ser	Glu	Glu	Tyr	Gln	Cys	Ala	
		515					520					525				
tgg	ctg	gat	atg	tat	cac	cgc	gtc	ttt	gat	cgc	gtc	agc	gcc	gtc	gtc	1632
Trp	Leu	Asp	Met	Tyr	His	Arg	Val	Phe	Asp	Arg	Val	Ser	Ala	Val	Val	
	530					535					540					
ggg	gaa	cag	gta	tgg	aat	ttc	gcc	gat	ttt	gcg	acc	tcg	caa	ggc	ata	1680
Gly	Glu	Gln	Val	Trp	Asn	Phe	Ala	Asp	Phe	Ala	Thr	Ser	Gln	Gly	Ile	
545					550					555					560	
ttg	cgc	gtt	ggc	ggg	aac	aag	aaa	ggg	atc	ttc	act	cgc	gac	cgc	aaa	1728
Leu	Arg	Val	Gly	Gly	Asn	Lys	Lys	Gly	Ile	Phe	Thr	Arg	Asp	Arg	Lys	
			565						570					575		
ccg	aag	tcg	gcg	gct	ttt	ctg	ctg	caa	aaa	cgc	tgg	act	ggc	atg	aac	1776
Pro	Lys	Ser	Ala	Ala	Phe	Leu	Leu	Gln	Lys	Arg	Trp	Thr	Gly	Met	Asn	
			580					585					590			
ttc	ggg	gaa	aaa	ccg	cag	cag	gga	ggc	aaa	caa	tga					1812
Phe	Gly	Glu	Lys	Pro	Gln	Gln	Gly	Gly	Lys	Gln	*					
		595					600									

<210> 106

<211> 603

<212> PRT

<213> Escherichia coli

<300>

<308> Genbank #S69414

<309> 1994-09-23

<400> 106

Met	Leu	Arg	Pro	Val	Glu	Thr	Pro	Thr	Arg	Glu	Ile	Lys	Lys	Leu	Asp	
1				5					10					15		
Gly	Leu	Trp	Ala	Phe	Ser	Leu	Asp	Arg	Glu	Asn	Cys	Gly	Ile	Asp	Gln	
			20					25					30			
Arg	Trp	Trp	Glu	Ser	Ala	Leu	Gln	Glu	Ser	Arg	Ala	Ile	Ala	Val	Pro	
		35					40					45				
Gly	Ser	Phe	Asn	Asp	Gln	Phe	Ala	Asp	Ala	Asp	Ile	Arg	Asn	Tyr	Ala	
	50				55						60					
Gly	Asn	Val	Trp	Tyr	Gln	Arg	Glu	Val	Phe	Ile	Pro	Lys	Gly	Trp	Ala	
65					70				75						80	
Gly	Gln	Arg	Ile	Val	Leu	Arg	Phe	Asp	Ala	Val	Thr	His	Tyr	Gly	Lys	
			85					90						95		
Val	Trp	Val	Asn	Asn	Gln	Glu	Val	Met	Glu	His	Gln	Gly	Gly	Tyr	Thr	
			100					105					110			
Pro	Phe	Glu	Ala	Asp	Val	Thr	Pro	Tyr	Val	Ile	Ala	Gly	Lys	Ser	Val	
		115					120					125				

-65-

Arg	Ile	Thr	Val	Cys	Val	Asn	Asn	Glu	Leu	Asn	Trp	Gln	Thr	Ile	Pro
130						135					140				
Pro	Gly	Met	Val	Ile	Thr	Asp	Glu	Asn	Gly	Lys	Lys	Lys	Gln	Ser	Tyr
145						150				155					160
Phe	His	Asp	Phe	Phe	Asn	Tyr	Ala	Gly	Ile	His	Arg	Ser	Val	Met	Leu
				165					170					175	
Tyr	Thr	Thr	Pro	Asn	Thr	Trp	Val	Asp	Asp	Ile	Thr	Val	Val	Thr	His
			180					185					190		
Val	Ala	Gln	Asp	Cys	Asn	His	Ala	Ser	Val	Asp	Trp	Gln	Val	Val	Ala
		195					200					205			
Asn	Gly	Asp	Val	Ser	Val	Glu	Leu	Arg	Asp	Ala	Asp	Gln	Gln	Val	Val
		210				215					220				
Ala	Thr	Gly	Gln	Gly	Thr	Ser	Gly	Thr	Leu	Gln	Val	Val	Asn	Pro	His
225						230				235					240
Leu	Trp	Gln	Pro	Gly	Glu	Gly	Tyr	Leu	Tyr	Glu	Leu	Cys	Val	Thr	Ala
				245					250					255	
Lys	Ser	Gln	Thr	Glu	Cys	Asp	Ile	Tyr	Pro	Leu	Arg	Val	Gly	Ile	Arg
			260					265					270		
Ser	Val	Ala	Val	Lys	Gly	Glu	Gln	Phe	Leu	Ile	Asn	His	Lys	Pro	Phe
		275					280					285			
Tyr	Phe	Thr	Gly	Phe	Gly	Arg	His	Glu	Asp	Ala	Asp	Leu	Arg	Gly	Lys
	290					295					300				
Gly	Phe	Asp	Asn	Val	Leu	Met	Val	His	Asp	His	Ala	Leu	Met	Asp	Trp
305					310					315					320
Ile	Gly	Ala	Asn	Ser	Tyr	Arg	Thr	Ser	His	Tyr	Pro	Tyr	Ala	Glu	Glu
			325						330					335	
Met	Leu	Asp	Trp	Ala	Asp	Glu	His	Gly	Ile	Val	Val	Ile	Asp	Glu	Thr
			340					345					350		
Ala	Ala	Val	Gly	Phe	Asn	Leu	Ser	Leu	Gly	Ile	Gly	Phe	Glu	Ala	Gly
		355					360					365			
Asn	Lys	Pro	Lys	Glu	Leu	Tyr	Ser	Glu	Glu	Ala	Val	Asn	Gly	Glu	Thr
	370					375					380				
Gln	Gln	Ala	His	Leu	Gln	Ala	Ile	Lys	Glu	Leu	Ile	Ala	Arg	Asp	Lys
385					390					395					400
Asn	His	Pro	Ser	Val	Val	Met	Trp	Ser	Ile	Ala	Asn	Glu	Pro	Asp	Thr
				405					410					415	
Arg	Pro	Gln	Gly	Ala	Arg	Glu	Tyr	Phe	Ala	Pro	Leu	Ala	Glu	Ala	Thr
			420					425					430		
Arg	Lys	Leu	Asp	Pro	Thr	Arg	Pro	Ile	Thr	Cys	Val	Asn	Val	Met	Phe
		435					440					445			
Cys	Asp	Ala	His	Thr	Asp	Thr	Ile	Ser	Asp	Leu	Phe	Asp	Val	Leu	Cys
	450					455					460				
Leu	Asn	Arg	Tyr	Tyr	Gly	Trp	Tyr	Val	Gln	Ser	Gly	Asp	Leu	Glu	Thr
465					470					475					480
Ala	Glu	Lys	Val	Leu	Glu	Lys	Glu	Leu	Leu	Ala	Trp	Gln	Glu	Lys	Leu
				485					490					495	
His	Gln	Pro	Ile	Ile	Ile	Thr	Glu	Tyr	Gly	Val	Asp	Thr	Leu	Ala	Gly
			500					505					510		
Leu	His	Ser	Met	Tyr	Thr	Asp	Met	Trp	Ser	Glu	Glu	Tyr	Gln	Cys	Ala
		515					520					525			
Trp	Leu	Asp	Met	Tyr	His	Arg	Val	Phe	Asp	Arg	Val	Ser	Ala	Val	Val
	530					535					540				
Gly	Glu	Gln	Val	Trp	Asn	Phe	Ala	Asp	Phe	Ala	Thr	Ser	Gln	Gly	Ile
545					550					555					560
Leu	Arg	Val	Gly	Gly	Asn	Lys	Lys	Gly	Ile	Phe	Thr	Arg	Asp	Arg	Lys
				565					570					575	
Pro	Lys	Ser	Ala	Ala	Phe	Leu	Leu	Gln	Lys	Arg	Trp	Thr	Gly	Met	Asn
			580					585					590		
Phe	Gly	Glu	Lys	Pro	Gln	Gln	Gly	Gly	Lys	Gln					
		595					600								

<210> 107

<211> 277

<212> DNA

<213> Artificial Sequence

-66-

<220>

<223> Nopaline Synthase Terminator Sequence

<300>

<308> U09365

<309> 1995-10-17

<400> 107

gagctcgaat	ttccccgatc	gttcaaacat	ttggcaataa	agtttcttaa	gattgaatcc	60
tggttgccggt	cttgcgatga	ttatcatata	atctctgttg	aattacgtta	agcatgtaat	120
aattaacatg	taatgcatga	cgttatttat	gagatgggtt	tttatgatta	gagtccccga	180
attatacatt	taatacgcga	tagaaaacaa	aatatagcgc	gcaaactagg	ataaattatc	240
gcgcgcgggtg	tcattctatgt	tactagatcg	ggaattc			277

<210> 108

<211> 3451

<212> DNA

<213> Artificial Sequence

<220>

<223> HindIII Fragment containing the beta-glucuronidase coding sequence, the rDNA intergenic spacer, and the Mast1 sequence

<400> 108

aagcttgacc	tggaatatcg	cgagtaaact	gaaaatcacg	gaaaatgaga	aatacacact	60
ttaggacgtg	aaatatggcg	aggaaaactg	aaaaagggtg	aaaatttaga	aatgtccact	120
gtaggacgtg	gaatatggca	agaaaactga	aaatcatgga	aaatgagaaa	catccacttg	180
acgacttgaa	aaatgacgaa	atcactaaaa	aacgtgaaaa	atgagaaatg	cacactgaag	240
gactccgcgg	gaattcgatt	gtgctagcca	atgtttaaca	agatgtcaag	cacaatgaat	300
gttggtggtt	ggtggtcgtg	gctggcgggtg	gtggaaaatt	gcgggtggttc	gagcggtagt	360
gatcggcgat	ggttggtggt	tgcagcgggtg	tttgatatcg	gaatcactta	tggtggttgt	420
cacaatggag	gtgcgtcatg	gttattgggtg	gttggtcatc	tatatatttt	tataataata	480
ttaagtattt	tacctatttt	ttacatatatt	tttattaaat	ttatgcattg	tttgtatttt	540
taaatagttt	ttatcgtact	tgttttataa	aatattttat	tatttttatgt	gttatattat	600
tacttgatgt	attggaaatt	ttctccattg	ttttttctat	atattataata	attttcttat	660
ttttttttgt	tttattatgt	atttttttcgt	ttttataata	atatttatta	aaaaaaatat	720
tattttttgt	aaatatatca	tttacaatgt	ttaaaagtca	tttgtgaata	tattagctaa	780
gttggtacttc	ttttgtgca	tttggtggtg	tacatgtcta	ttatgattct	ctggccaaaa	840
catgtctact	cctgtcactt	gggttttttt	ttttaagaca	taatcactag	tgatttatatc	900
tagactgaag	gcgggaaacg	acaatctgat	catgagcggg	gaattaaggg	agtcacgtta	960
tgacccccgc	cgatgacgcg	ggacaagccg	ttttacgttt	ggaactgaca	gaaccgcaac	1020
gttgaaggag	ccactcagcc	gcgggtttct	ggagtttaat	gagctaagca	catacgtcag	1080
aaaccattat	tgcgcgttca	aaagtgcgct	aagggtcacta	tcagctagca	aatattttctt	1140
gtcaaaaatg	ctccactgac	gttccataaa	ttccctctcg	tatccaatta	gagtctcata	1200
ttcactctca	atccaaataa	tctgcaccgg	atctcgagat	cgaattcccc	cggcgcgcaa	1260
ttcactagtg	gatccccggg	tacggtcagt	cccttatgtt	acgtcctgta	gaaaccccaa	1320
cccgtagaat	caaaaaactc	gacggcctgt	gggcattcag	tctggatcgc	gaaaaactgtg	1380
gaattgagca	gcgttggtgg	gaaagcgcgt	tacaagaaag	ccgggcaatt	gctgtgccag	1440
gcagttttta	cgatcagttc	gccgatgcag	atattcgtaa	ttatgtgggc	aacgtctggg	1500
atcagcgcga	agtctttata	ccgaaagggt	gggcaggcca	gcgtatcgtg	ctgcgtttcg	1560
atgcggtcac	tcattacggc	aaagtgtggg	tcaataatca	ggaagtgatg	gagcatcagg	1620
gcggctatac	gccatttgaa	gccgatgtca	cgccgtatgt	tattgccggg	aaaagtgtac	1680
gtatcacagt	ttgtgtgaac	aacgaactga	actggcagac	tatccgcggc	ggaatggtga	1740
ttaccgacga	aaacggcaag	aaaaagcagt	cttacttcca	tgatttcttt	aactacggcg	1800
ggatccatcg	cagcgtaatg	ctctacacca	cgccgaacac	ctgggtggac	gatatacccg	1860
tggtgacgca	tgctgcgcaa	gactgtaacc	acgcgtctgt	tgactggcag	gtggtggcca	1920
atggtgatgt	cagcgttgaa	ctgcgtgatg	cggtacaaca	ggtggttgca	actggacaag	1980
gcaccagcgg	gactttgcaa	gtggtgaatc	cgcacctctg	gcaaccgggt	gaagggtatc	2040
tctatgaact	gtacgtcaca	gccaaaagcc	agacagagtg	tgatatctac	ccgctgcgcg	2100
tcggcatccg	gtcagtgcca	gtgaaggggc	aacagttcct	gtaccaaccac	aaaccggtct	2160
actttactgg	ctttggccgt	catgaagatg	cggatttgcg	cggcaaaggga	ttcgataacg	2220
tgcgtatggt	gcacgatcac	gcattaatgg	actggattgg	ggccaactcc	taccgtacct	2280
cgctattacc	ttacgctgaa	gagatgctcg	actgggcaga	tgaacatggc	atcgtgggtga	2340
ttgatgaaac	tgcagctgtc	ggctttaacc	tctctttagg	cattgggttc	gaagcgggca	2400
acaagccgaa	agaactgtac	agcgaagagg	cagtcaacgg	ggaaactcag	caggcgcact	2460
tacaggcgat	taaagagctg	atagcgcgtg	acaaaaacca	cccaagcgtg	gtgatgtgga	2520

-67-

gtattgccaa	cgaaccggat	acccgtccgc	aagggtgcacg	ggaatatttc	gcgccactgg	2580
cggaagcaac	gcgtaaactc	gatccgacgc	gtccgatcac	ctgcgtcaat	gtaatgttct	2640
gcgacgctca	caccgatacc	atcagcgatc	tctttgatgt	gctgtgcctg	aaccgttatt	2700
acgggttggt	tgtccaaagc	ggcgatttgg	aaacggcaga	gaaggtagctg	gaaaaagaac	2760
ttctggcctg	gcaggagaaa	ctgcatcagc	cgattatcat	caccgaatac	ggcgtggata	2820
cgtttagccgg	gctgcactca	atgtacaccg	acatgtggag	tgaagagtat	cagtgtgcat	2880
ggctggatat	gtatcaccgc	gtctttgatc	gcgtcagcgc	cgctcgtcgt	gaacagggtat	2940
ggaatttcgc	cgattttgcg	acctcgcaag	gcatattgcg	cgttggcggg	aacaagaagg	3000
ggatcttcac	ccgcgaccgc	aaaccgaaat	cgccggcctt	tctgctgcaa	aaacgctgga	3060
ctggcatgaa	cttcggtgaa	aaaccgcagc	agggaggcaa	acaatgaatc	aacaactctc	3120
ctggcgaccc	atcgctgggt	acagcctcgg	gaattgcgta	ccgagctcga	atttccccga	3180
tcgttcaaac	atttggcaat	aaagtttctt	aagattgaa	cctgttgccg	gtcttgcgat	3240
gattatcata	taatttctgt	tgaattacgt	taagcatgta	ataattaaca	tgtaatgcat	3300
gacgttattt	atgagatggg	tttttatgat	tagagtcctc	caattataca	tttaatacgc	3360
gatagaaaac	aaaaatatgc	gcgcaaatca	ggataaatta	tcgcgcgcgg	tgcatctat	3420
gttactagat	cggaattcgc	atatcaagct	t			3451

<210> 109

<211> 14627

<212> DNA

<213> Artificial Sequence

<220>

<223> pAg11a Plasmid

<400> 109

catgccaaac	acagggttcc	cctcggggtc	aaagtacttt	gatccaaccc	ctccgctgct	60
atagtgacgt	cggcttctga	cgttcagtcg	agccgtcttc	tgaaaacgac	atgtcgcaca	120
agtcctaagt	tacgcgacag	gctgccgccc	tgcccttttc	ctggcgcttt	cttgtcgcgt	180
gttttagtcg	cataaagtag	aatacttgcc	actagaaccg	gagacattac	gccatgaaca	240
agagcgccgc	cgctggcctg	ctgggctatg	cccgcgtcag	caccgacgac	caggacttga	300
ccaaccaacg	ggccgaaactg	cacgcggcgc	gctgcaccaa	gctgttttcc	gagaagatca	360
ccggcaccag	gcgcgaccgc	ccggagctgg	ccaggatgct	tgaccaccta	cgccctggcg	420
acgtttgtgac	agtgaccagg	ctagaccgcc	tgcccccgcg	cacccgcgac	ctactggaca	480
ttgccgagcg	catccaggag	gcccggcgcg	gcctgcgtag	cctggcagag	ccgtgggccc	540
acaccaccac	gcccggccgc	cgcatggtgt	tgaccgtgtt	cgccggcatt	gccgagttcg	600
agcgttccct	aatcatcgac	cgaccccgga	gcggggcgca	ggccgcgcaag	gcccgaggcg	660
tgaagtttgg	cccccgccct	accctcaccc	cggcacagat	cgcgacgccc	cgcgagctga	720
tcgaccagga	aggcgcgacc	gtgaaaagag	cggtgcact	gcttggcgtg	catcgctcga	780
ccctgtaccg	cgcaacttag	cgacgcgagg	aagtgcagcc	caccgaggcc	aggcgggcgc	840
gtgccttccg	tgaggacgca	ttgaccgagg	ccgacgccct	ggcgcccgcc	gagaatgaac	900
gccaagagga	acaagcatga	aaccgcacca	ggcagggccag	gacgaaccgt	ttttcattac	960
cgaagagatc	gaggcggaga	tgatcgcgcc	cggttacgtg	ttcgagccgc	ccgcgcacgt	1020
ctcaaccgtg	cggctgcatg	aaatcctggc	cggtttgtct	gatgccaaag	tgggcgccctg	1080
gccggccagc	ttggccgctg	aagaaaaccga	gcgcgcgctg	ctaaaaaggt	gatgtgtatt	1140
tgagtataaac	agcttgcgtc	atgcggtcgc	tgcttatatg	atgcgatgag	taataaaca	1200
aatacgcaag	gggaacgcat	gaaggttatc	gctgtactta	accagaaagg	cgggtcaggc	1260
aagacgacca	tcgcaaccca	tctagcccg	gccttgcaac	tcgcccgggg	cgatgttctg	1320
ttagtcgatt	ccgatcccca	gggcagtgcc	cgcgattggg	cgcccgctgc	ggaagatcaa	1380
ccgctaaccg	ttgtcggcat	cgaccgccc	acgattgacc	gcgacgtgaa	ggccatcgcc	1440
cggcgcgact	tcgtagtgat	cgacggagcg	cccaggcg	cggaactggc	tgtgtccgcg	1500
atcaaggcag	ccgacttcgt	gctgattccg	gtgcagccaa	gcccttacga	catatggg	1560
accgcccagc	tggtggagct	ggttaagcag	cgcattgagg	tcacggatgg	aaggctacaa	1620
gcgccctttg	tcgtgtcgcg	ggcgatcaaa	ggcacgcgca	tcggcggtga	ggttgccgag	1680
gcgctggccg	ggtacgagct	gcccattctt	gagtcgccga	tcacgcagcg	cgtgagctac	1740
ccaggcactg	ccgcgcgccc	cacaaccggt	cttgaatcag	aaccggagg	cgacgctgcc	1800
cgcgaggtcc	aggcgcgtgg	cgctgaaatt	aaatcaaac	tcatttgagt	taatgaggt	1860
aagagaaaat	gagcaaaaag	acaaacacgc	taagtgcgg	ccgtccgagc	gcacgcagca	1920
gcaaggctgc	aacgttggcc	agcctggcag	acacgccagc	catgaagcgg	gtcaactttc	1980
agttgcccgc	ggaagatcac	accaagctga	agatgtacgc	ggtacgcaa	ggcaagacca	2040
ttaccgagct	gctatctgaa	tacatcgcg	agtcaccaga	gtaaatgagc	aaatgaataa	2100
atgagtagat	gaatttttag	ggctaaagga	ggcgcatgg	aaaatcaaga	acaaccaggc	2160
accgacgccc	tggaatgccc	catgtgtgga	ggaacgggg	gttggccagg	cgtaagcgcc	2220
tgggttctct	ccatggcctg	caatggcact	ggaaccccca	agcccgagga	atcggcgtga	2280
cggtcgcaaa	ccatccggcc	cggtacaaat	cggcgcggcg	ctgggtgatg	acctgggtga	2340
gaagttgaag	gcccgcgagg	ccgcccagcg	gcaacgcac	gaggcagaag	cacgccccgg	2400
tgaatcgtgg	caagcggccg	ctgatcgaat	ccgcaaaaga	tcccggcaac	cgccggcagc	2460

-68-

cgggtgcgcgcg	tgcattagga	agccgccccaa	gggcgacgag	caaccagatt	ttttcgttcc	2520
gatgctctat	gacgtgggca	cccgcgatag	tccgagcatc	atggacgtgg	ccgttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	gggtgatccgc	tacgagcttc	cagacgggca	2640
cgtagagggtt	tccgcagggc	cggccggcat	ggccagtgtg	tgggattacg	acctgggtact	2700
gatggcgggtt	tcccatctaa	ccgaatccat	gaaccgatac	cggaaggga	agggagacaa	2760
gcccggccgc	gtgttccgtc	cacacgttgc	ggacgtactc	aagttctgcc	ggcgagccga	2820
tggcggaaaag	cagaaagacg	acctggtaga	aacctgcatt	cggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cggcgccctg	gtgacggtat	ccgaggggta	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaaacc	gggcgggccgg	agtacatcga	3000
gatcgagcta	gctgattgga	tgtaccgcga	gatcacagaa	ggcaagaacc	cggacgtgct	3060
gacggttcac	cccgattact	ttttgatcga	tcccgccatc	ggccggtttc	tctaccgctc	3120
ggcacgcccgc	gcccagggca	aggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
cagtggcagc	gcccggagagt	tcaagaagtt	ctgtttcacc	gtgcgcgaagc	tgatcgggtc	3240
aatgacctg	ccggagtacg	atttgaagga	cggaggcggg	caggctggcc	cgatccctagt	3300
catgcgctac	cgcaacctga	tgcaggggca	agcatccgcc	ggttccctaat	gtacggagca	3360
gatgtcatggg	caaatggccc	tagcagggga	aaaagggtcga	aaaggctctct	ttcctgtgga	3420
tagcacgtac	attgggaacc	caaagccgta	cattgggaac	cggaaccctg	acattgggga	3480
cccaaagccg	tacattggga	accgggtcac	catgtaagtg	actgatataa	aagagaaaaa	3540
aggcgattttt	tccgcctaaa	actctttaaa	acttattaaa	actcttaaaa	cccgcctggc	3600
ctgtgcataa	ctgtctggcc	agcgcacagc	cgaagagctg	caaaaagcgc	ctaccgctcg	3660
gtcgctgcgc	tccctacgcc	ccgcgccttc	gcgtcgccct	atcgccggccg	ctggccgctc	3720
aaaaatgggt	ggcctacggc	caggcaatct	accaggggcg	ggacaagccg	cgcctgcgcc	3780
actcgaccgc	cgccgcccac	atcaaggcac	cctgcctcgc	gcgtttcggg	gatgacgggtg	3840
aaaacctctg	acacatgcag	ctcccggaag	cggtcacagc	ttgtctgtaa	gccgatggcg	3900
ggagcagaca	agcccgtcag	ggcgcgtcag	cgggtgttgg	cgggtgtcgg	ggcgcagcca	3960
tgaccacgtc	acgtagcgat	agcggagtgt	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgcacc	atatgcgggt	tgaatatccg	cacagatgcg	taaggagaaa	4080
ataccgcate	aggcgctctt	ccgcttccct	gtcactgac	tccgtgcgct	cgggtcgttcg	4140
gctgcggcga	gcgggtatcag	ctcactcaaa	ggcggttaata	cggttatcca	cagaatcagg	4200
ggataacgca	ggaaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	4260
ggccgcggtt	ctggcggtttt	tccataggct	ccgccccctt	gacgagcate	acaaaaatcg	4320
acgctcaagt	cagagggtggc	gaaacccgac	aggactataa	agataccagg	cgtttccccc	4380
tgggaagctcc	ctcgtgcgct	ctcctgttcc	gaccttgcgc	cttaccggat	acctgtccgc	4440
ctttctccct	tccgggaagcg	tggcgctttc	tcatagctca	cgctgtagggt	atctcagttc	4500
gggtgtaggtc	gttcgctcca	agctgggctg	tgtgcacgaa	ccccccggtc	agccccagccg	4560
ctgcgcctta	tccggtaact	atcgtcttga	gtccaaaccg	gtaagacacg	acttatcgcc	4620
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	4680
gttcttgaag	tggtggccta	actacggcta	cactagaagg	acagtatttg	gtatctgcgc	4740
tctgctgaag	ccagttacct	tccgaaaaag	agttggtagc	tcttgatccg	gcaaacacaa	4800
caccgctgggt	agcgggtgggt	tttttggttg	caagcagcag	attacgcgca	gaaaaaaagg	4860
atctcaagaa	gatcctttga	tctttcttac	ggggtctgac	gtcagtgga	acgaaaaactc	4920
acgttaagggt	atttttgta	tgcattctag	gtactaaaaa	aattcatcca	gtaaaaatata	4980
atattttatt	ttctcccaat	caggcttgat	cccagtaag	tcaaaaaata	gtcgcacata	5040
ctgttcttcc	ccgatattct	ccctgatcga	ccggacgcag	aaggcaatgt	cataccactt	5100
gtccgccttg	ccgcttctcc	caagatcaat	aaagccactt	actttgccat	ctttcacaaa	5160
gatgttgctg	tctcccagggt	cgccgtggga	aaagacaagt	tccctcttcgg	gcttttccgt	5220
ctttaaaaaa	tcatacagct	cgcgcgggat	tttaaatgga	gtgtcttctt	cccagttttc	5280
gcaatccaca	tccggccagat	cgttattcag	taagtaatcc	aattcggcta	agcggctgtc	5340
taagctattc	gtatagggac	aatccgatat	gtcgatggag	tgaagagacc	tgatgcactc	5400
cgcatacagc	tcgataatct	tttcagggtt	ttgttcatct	tcataactct	ccgagcaaa	5460
gacgccatcg	gectcactca	tgagcagatt	gtccagccca	tcatgcccgtt	caaagtgcag	5520
gacctttgga	acaggcagct	ttccttccag	ccatagcatc	atgtcctttt	cccgttccac	5580
atcatagggtg	gtccctttat	accggctgtc	cgctcattttt	aaatatagggt	tttcattttc	5640
tcccaccagc	ttatatacct	tagcaggaga	cattccttcc	gtatctttta	cgcagcggta	5700
tttttcgatac	agttttttca	attccgggtga	tattctcatt	ttagccattt	attatttctc	5760
tccctttttc	tacagtattt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
aattcactgt	tcccttgcat	ctaaaacctt	aaataccaga	aaacagcttt	ttcaaagttg	5880
ttttcaaaagt	tggcgataaa	catagtatcg	acggagccga	ttttgaaacc	gcgggtgatca	5940
caggcagcaa	cgctctgtca	tcgttacaat	caacatgcta	ccctccgcga	gatcatccgt	6000
gtttcaaac	cggcagctta	gttgccgttc	ttccgaatag	catcggtaac	atgagcaag	6060
tctgcccgtc	tacaacggct	ctcccgtgga	cgccgtcccg	gactgatggg	ctgctgtgat	6120
cgagtgggtg	ttttgtgccc	agctgcccgt	cggggagctg	ttggctgggt	gggtggcagga	6180
tatatgtgtg	ttgaaacaaa	ttgacgctta	gacaacctaa	taacacattg	cggagctttt	6240
taatgtactg	aattaacgcc	gaattaatc	gggggatctg	gatttttagta	ctggattttg	6300
gtttttaggaa	ttagaaatct	tattgataga	agttattttac	aaatacaaat	acataactaag	6360
ggtttcttat	atgctcaaca	catgagcgaa	accctatagg	aaccctaatt	cccttatctg	6420
ggaactactc	acacattatt	atggagaaac	tcgagtcaaa	tctcggtgac	gggcaggacc	6480

ggacggggcg	gtaccggcag	gctgaagtcc	agctgccaga	aacccacgtc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgctcgtgce	6600
atgcgacgc	tcgggtcggt	gggcagcccg	atgacagcga	ccacgctctt	gaagccctgt	6660
gcctccaggg	acttcagcag	gtgggtgtag	agcgtggagc	ccagtcccgt	ccgctgggtg	6720
cggggggaga	cgtacacggg	cgactcggcc	gtccagtcgt	aggcgttgcg	tgccttccag	6780
gggcccgcgt	aggcagtgcc	ggcgacctcg	ccgtccacct	cggcgacgag	ccagggatag	6840
cgctcccgc	gacggacgag	gtcgtccgtc	cactcctgcg	gttcctgctg	ctcggtagcg	6900
aagttgaccg	tgcttgtctc	gatgtagtgg	ttgacgatgg	tgacagccgc	cggcagtgcc	6960
ggctcggttg	cacggcggtg	gtcggccggg	cgctggtctg	ggctcatggg	agactcgaga	7020
gagatagatt	tgtagagaga	gactgggtgat	ttcagcgtgt	cctctccaaa	tgaatgaac	7080
ttccttatat	agaggaaggt	cttgcggaagg	atagtgggat	tgtgcgtcat	cccttacgtc	7140
agtggagata	tcacatcaat	ccacttgctt	tgaagacgtg	gttggaaacgt	cttctttttc	7200
cacgatgctc	ctcgtgggtg	gggggtccatc	tttgggacca	ctgtcggcag	aggcatcttg	7260
aacgatagcc	tttcttttat	cgcaatgatg	gcattttag	gtgccacctt	ccttttctac	7320
tgctcttttg	atgaagtgaac	agatagctgg	cgatggaaat	ccgaggaggt	ttcccgatat	7380
taccctttgt	tgaaaagtct	caatagccct	ttgggtcttct	gagactgtat	ctttgatatt	7440
cttgagtag	acgagagtgt	cgtgctccac	catgttatca	catcaatcca	cttgctttga	7500
agacgtgggt	ggaaacgtctt	ctttttccac	gatgctcctc	gtgggtgggg	gtccatcttt	7560
gggaccactg	tcggcagagg	catcttgaac	gatagccttt	cctttatcgc	aatgatggca	7620
tttgtaggtg	ccaccttccct	tttctactgt	ccttttgatg	aagtgcacaga	tagctgggga	7680
atggaatccg	aggaggtttc	ccgatattac	cctttgttga	aaagtctcaa	tagccctttg	7740
gtcttctgag	actgtatctt	tgatattctt	ggagtagacg	agagtgtcgt	gtccaccat	7800
gttggcaagc	tgctctagcc	aatacgcaaa	ccgcctctcc	ccgcgcgttg	gccgattcat	7860
taatgcagct	ggcacgacag	gtttcccgac	tggaagcgg	gcagtgcgag	caacgcaatt	7920
aatgtgagtt	agctcactca	ttaggcacc	caggctttac	actttatgct	tccggctcgt	7980
atgttgtgtg	gaattgtgag	cggataacaa	tttcacacag	gaaacagcta	tgaccatgat	8040
tacgaattcg	agccttgact	agagggtcga	cggtatacag	acatgataag	atacattgat	8100
gagtttgagc	aaaccacaac	tagaatgcag	tgaaaaaaat	gctttatttg	tgaattttgt	8160
gatgtatttg	ctttattttgt	aaccattata	agctgcaata	aacaagttag	gggtggcgaa	8220
gaactccagc	atgagatccc	cgcgtggag	gatcatccag	ccggcgtccc	ggaaaaacgt	8280
tccgaagccc	aacctttcat	agaaggcggc	ggtggaatcg	aaatctcgtg	gcacgtgtca	8340
gtcctgtccc	tcggccacga	agtgcacgca	gttgccggcc	gggtcgcgca	gggcgaactc	8400
ccgcccccc	ggctgtcgc	cgatctcggt	catggcggc	ccggaggcgt	ccgggaactc	8460
cgtggacacg	acctccgacc	actcggcgta	cagctcgtcc	aggccgcgca	cccacacca	8520
ggccagggtg	ttgtccggca	ccacctgggtc	ctggaccgcy	ctgatgaaca	gggtcacgtc	8580
gtcccggacc	acaccggcga	agtcgtcctc	cacgaagtcc	cgggagaacc	cgagccggtc	8640
gggtccagaac	tcgaccgctc	cggcgacgtc	gcgcgcgggtg	agcaccggaa	cggcactggg	8700
caacttggcc	atggatccag	atttcgctca	agttagtata	aaaaagcagg	cttcaatcct	8760
gcaggaattc	tattgagact	ctcgtctact	ccaagaatat	caaagataca	gtctcagaag	8820
accaaagggc	tattcagact	tttcaacaaa	gggtaatatc	gggaaacctc	ctcggattcc	8880
attgcccagc	tatctgtcac	ttcatcaaaa	ggacagtaga	aaagggaagg	ggcacctaca	8940
aatgccatca	ttgcgataaa	ggaaaggcta	tcgttcaaga	tgctctgccc	gacagtgggt	9000
ccaaagatgg	acccccaccc	acgaggagca	tcgtggaaaa	agaagacgtt	ccaaccacgt	9060
cttcaaagca	agtggtatga	tgtgataaca	tggtggagca	cgacactctc	gtctactcca	9120
agaatatcaa	agatacagtc	tcagaagacc	aaagggtat	tgagactttt	caacaaaggg	9180
taatatcggt	aaacctcctc	ggattccatt	gcccagctat	ctgtcacttc	atcaaaagga	9240
cagtagaaaa	ggaaggtggc	acctacaaat	gccatcattg	cgataaagga	aaggctatcg	9300
ttcaagatgc	ctctgccgac	agtgggtccc	aagatggacc	cccaccacag	aggagcatcg	9360
tggaaaaaaga	agacgttcca	accacgtctt	caaagcaagt	ggattgatgt	gatatctcca	9420
ctgacgtaag	ggatgacgca	caatcccact	atccttcgca	agaccttcct	ctatataagg	9480
aagttcattt	catttgagga	ggacacgctg	aatcaccag	tctctctcta	caaactctac	9540
tctctcgagc	tttcgcagat	ccgggggggg	aatgagatat	gaaaaagcct	gaactcaccg	9600
cgacgtctgt	cgagaagttt	ctgatcgaaa	agttcgacag	cgtctccgac	ctgatgcagc	9660
tctcggaggg	cgaagaatct	cgtgctttca	gcttcgatgt	aggaggcgt	ggatatgtcc	9720
tgccggtaaa	tagctgcgcc	gatgggtttct	acaaagatcg	ttatgtttat	cggcactttg	9780
catcggccgc	gctcccgatt	ccggaagtgc	ttgacattgg	ggagtttagc	gagagcctga	9840
cctattgcat	ctcccccgct	gcacaggggtg	tcactgttga	agacctgcct	gaaaccgaac	9900
tgcctcgctg	tctacaaccg	gtcgcggagg	catggtatgc	gatcgctgcy	gccgatctta	9960
gccagacgag	cgggttcggc	ccattcggac	cgcaagggaat	cgggtcaatac	actacatggc	10020
gtgatttcat	tgccgcgatt	gctgatcccc	atgtgtatca	ctggcaaaact	gtgatggagc	10080
acaccgtcag	tgctccgtc	gcgcaggctc	ggttggtctg	gatgcttttg	gccgaggact	10140
gccccgaagt	ccggcacctc	gtgcacgcgg	atttcggctc	caacaatgtc	ctgacggaca	10200
atggccgcat	aacagcgggtc	attgactgga	gcgaggcgat	gttcggggat	tcccaatagc	10260
aggtcgccaa	catcttcttc	tggaggccgt	ggttggtctg	tatggagcag	cagacgcgct	10320
acttcgagcg	gaggcatccg	gagcttgacg	gatcgccacg	actccgggcy	tatatgctcc	10380
gcattgggtct	tgaccaactc	tatcagagct	tggttgacgg	caatttcgat	gatgcagctt	10440
gggcgcaggg	tcgatgcgac	gcaatcgtcc	gatccggagc	cgggactgtc	gggcgtacac	10500

aaatcgcccc	cagaagcgcg	gccgtctgga	ccgatggctg	tgtagaagta	ctcgccgata	10560
gtggaaaccg	acgccccagc	actcgtccga	gggcaaagaa	atagagtaga	tgccgaccgg	10620
atctgtcgat	cgacaagctc	gagtttctcc	ataataatgt	gtgagtagtt	cccagataag	10680
ggaattaggg	ttcctatagg	gtttcgctca	tgtgttgagc	atataagaaa	cccttagtat	10740
gtattttgat	ttgtaaaaaa	cttctatcaa	taaaatttct	aattcctaaa	acaaaaatcc	10800
agtactaaaa	tccagatccc	ccgaattaat	tcggcggttaa	ttcagatcaa	gcttgacctg	10860
gaatatcgcg	agtaaaactga	aaatcacgga	aaatgagaaa	tacacacttt	aggacgtgaa	10920
atatggcgag	gaaaaactgaa	aaagggtggaa	aatttagaaa	tgtccactgt	aggacgtgga	10980
atatggcaag	aaaaactgaaa	atcatggaaa	atgagaaaca	tccacttgac	gacttgaaaa	11040
atgacgaaat	cactaaaaaaa	cgtgaaaaat	gagaaatgca	cactgaagga	ctccgcggga	11100
attcgattgt	gctagccaat	gtttaacaag	atgtcaagca	caatgaatgt	tgggtggttg	11160
tggctcgtggc	tggcggtgggt	ggaaaaattgc	gggtggttcga	gcggtagtga	tcggcgatgg	11220
ttgggtgtttg	cagcggtgtt	tgatatcgga	atcacttatg	gtggttgtca	caatggaggt	11280
gcgtcatggt	tattgggtggt	tggctatcta	tatatattta	taataatatt	aagtatttta	11340
cctattttttt	acatattttt	tattaaattt	atgcattggt	tgtattttta	aatagttttt	11400
atcgtactttg	ttttataaaa	tatttttatta	ttttatgtgt	tatatattta	ctttagtatt	11460
tggaaatttt	tttctatat	tttctatat	ttataaata	tttctattt	tttctgttt	11520
tattatgtat	tttttcgttt	tataataaat	atttattaaa	aaaaatatta	tttttgtaaa	11580
atatatcatt	tacaatgttt	aaaagtcat	tgtgaatata	ttagctaaagt	tgtacttctt	11640
tttgtgcatt	tgggtgtgta	catgtctatt	atgattctct	ggccaaaaca	tgtctactcc	11700
tgtcacttgg	gttttttttt	ttaaagacata	atcactagt	attatatcta	tactgaaggg	11760
gggaaacgac	aatctgatca	tgagcggaga	attaagggag	tcacgttatg	accccgcccg	11820
atgacgcggg	acaagccgtt	ttacgtttgg	aactgacaga	accgcaacgt	tgaaggagcc	11880
actcagccgc	gggtttctgg	agtttaatga	gctaagcaca	tacgtcagaa	accattattg	11940
cgcggtcaaa	agtcgcctaa	ggtcactatc	agctagcaaa	tatttcttgt	caaaaatgct	12000
ccactgacgt	tccataaatt	cccctcggta	tccaattaga	gtctcatatt	cactctcaat	12060
ccaaataatc	tgcaccggat	ctcgagatcg	aattcccgcg	gccgcgaatt	cactagtggg	12120
tccccgggta	cggtcagtc	cttatgttac	gtcctgtaga	aaccccaacc	cgtgaaatca	12180
aaaaactcga	cggcctgtgg	gcattcagtc	tggatcgga	aaactgtgga	attgagcagc	12240
gttgggtggga	aagcgcgtta	caagaaagcc	gggcaattgc	tgtgccaggg	agttttaacg	12300
atcagttcgc	cgatgcagat	attcgttaatt	atgtgggcaa	cgtctgggtat	cagcgcgaag	12360
tctttatacc	gaaaggttgg	gcaggccagc	gtatcgtgct	gcgtttcgat	gcggtcactc	12420
aattacggcaa	agtggtgggtc	aataatcagg	aagtgtatga	gcatacaggg	aggtatcgc	12480
catttgaagc	cgatgtcacg	ccgtatgtta	ttgccgggaa	aagtgtacgt	atcacagttt	12540
gtgtgaacaa	cgaactgaac	tggcagacta	tcccgcgggg	aatgggtgatt	accgacgaaa	12600
acggcaagaa	aaagcagtc	tacttccatg	atttctttaa	ctacgcgggg	atccatcgga	12660
gcgtaatgct	ctacaccacg	ccgaacacct	gggtggacga	tatcacccgtg	gtgacgcatg	12720
tccgcgaaga	ctgtaaccac	gcgtctgttg	actggcaggt	gggtggccaat	gggtgatgtca	12780
gcgttgaact	gctgatgagc	gatcaaacgg	tggttgcaac	tggacaaggg	accagcggga	12840
ctttgcaagt	ggtgaatccg	cacctctggc	aaccgggtga	aggttatctc	tatgaactgt	12900
acgtcacagc	caaaaagccag	acagagtggtg	atatctaccc	gctgcgcgtc	ggcatccggt	12960
cagtgccagt	gaagggcgaa	cagttcctga	tcaaccacaa	accgttctac	tttactggct	13020
ttggccgtca	tgaagatgag	gatttgccgc	gcaaaggatt	cgataaacgtg	ctgatgggtc	13080
acgatcacgc	attaatggac	tggattgggg	ccaactccta	ccgtacctcg	cattaccctt	13140
acgctgaaga	gatgctcgac	tgggcagatg	aacatggcat	cgtgggtgatt	gatgaaactg	13200
cagctgtcgg	ctttaacctc	tcttttaggca	ttggtttcga	agcgggcaac	aagccgaaa	13260
aactgtacag	cgaagaggca	gtcaacgggg	aaactcagca	ggcgcaacta	caggcgatta	13320
aagagctgat	agcgcgtgac	aaaaaccacc	caagcgtggg	gatgtggagt	attgccaacg	13380
aaccggatac	ccgtccgcaa	gggtgcacggg	aatatttcgc	gccactggcg	gaagcaacgc	13440
gtaaactcga	tccgacgcgt	ccgatcacct	gcgtcaatgt	aatgttctgc	gacgtccaca	13500
ccgataccat	cagcgatctc	tttgatgtgc	tgtgcctgaa	ccgttattac	ggttggtagt	13560
tccaaagcgg	cgatttgga	acggcagaga	aggatctgga	aaaagaactt	ctggcctggc	13620
aggagaaact	gcatcagccg	attatcatca	ccgaatacgg	cgtggatac	ttagccgggg	13680
tgcactcaat	gtacaccgac	atgtggagtg	aagagtatca	gtgtgcattg	ctggatattg	13740
atcacccgct	ctttgatcgc	gtcagcgccg	tcgtcgggtg	acaggtatgg	aatttcgccc	13800
attttgcgac	ctcgcaaggc	atattcgcgc	ttggcggtaa	caagaagggg	atcttcaccc	13860
gcgaccgcaa	accgaagtcg	gcggttttct	tgtgcacaaa	acgtggact	ggcatgaact	13920
tcgggtgaaaa	accgcagcag	ggaggcaaac	aatgaatcaa	caactctcct	ggcgacacat	13980
cgtcgggtac	agcctcgggg	attgcgtacc	gagctcgaat	ttccccgatc	gttcaaacat	14040
ttggcaataa	agtttcttaa	agtgaaatcc	tgttgccggt	cttgcgatga	ttatcatata	14100
atttctgttg	aatcagttta	agcatgtaat	aatcaacatg	taatgcattg	cgttatttat	14160
gagatgggtt	tttatgatta	gagtcocgca	attatacatt	taatacgcga	tagaaaaaaa	14220
aatatagcgc	gcaaactagg	ataaattatc	gcgcgcgggt	tcactatgt	tactagatcg	14280
ggaattcgat	atcaagcttg	gcactggccg	tcgttttaca	acgtcgtgac	tgggaaaaac	14340
ctggcggttac	ccaacttaat	cgccttgcag	cacatcccc	tttcgccagc	tggcgtaata	14400
gcgaagaggg	ccgcaccgat	cgcccttccc	aacagttgcg	cagcctgaat	ggcgaaatgct	14460
agagcagctt	gagcttggat	cagattgtcg	tttcccgcct	tcagttttaa	ctatcagttg	14520

-71-

ttgacaggat atattggcgg gtaaacctaa gagaaaagag cgtttattag aataacggat 14580
atttaaaagg gcgtgaaaag gtttatccgt tcgtccattt gtatgtg 14627

<210> 110

<211> 9080

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeo (6XHS4) 2eGFP Plasmid

<400> 110

cagttgcccgg ccgggtcgcg cagggcggaac tccccccccc acggetgctc gccgatctcg 60
gtcatggccgg gcccgaggcg gtcccgggaag ttccgtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggccgcg caccacacacc caggccaggg tgttgtccgg caccacctgg 180
tcctggaccg cgctgatgaa cagggtcacg tcgtcccggg ccacaccggc gaagtctctc 240
tccacgaagt cccgggagaa cccgagccgg tcgggtccaga actcgaccgc tccggcgacg 300
tcgcgcgcgg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgatatc gaattcctgc 420
agcccccggg atccgctcac ggggacagcc cccccccaaa gccccagggg atgtaattac 480
gtccctcccc cgctaggggg cagcagcgag ccgcccgggg ctccgctccg gtccggcgct 540
ccccccgcag ccccgagccg gcagcgtgcg gggacagccc gggcacgggg aaggtggcac 600
gggatcgctt tcctctgaac gcttctcgct gctctttgag cctgcagaca cctgggggat 660
acggggccgc ggatccgctc acggggacag cccccccca aagccccag ggatgtaatt 720
acgtccctcc cccgctaggg ggcagcagcg agccgcccgg ggctccgctc cggctccggcg 780
ctccccccgc atccccgagc cggcagcgtg cggggacagc ccggggcacg ggaaggtggc 840
acgggatcgc tttcctctga acgcttctcg ctgctctttg agcctgcaga cacctggggg 900
atacggggcc gcggatccgc tcacggggac agcccccccc caaagcccc agggatgtaa 960
ttacgtccct cccccgctag ggggcagcag cgagccgccc gggggtccgc tccgggtccg 1020
cgctcccccc gcctccccga gccgggagac tgccgggaca gcccgggcac ggggaagggt 1080
gcacggggatc gctttcctct gaacgcttct cgctgctctt tgagcctgca gacacctggg 1140
ggatacgggg ccgcggaatcc gctcacgggg acagccccc cccaaagccc ccagggatgt 1200
aattacgtcc ctccccgctc agggggcagc agcagccgc cgggggctcc gctccggtcc 1260
ggcgctcccc ccgcctcccc gagccggcag cgtgcgggga cagcccgggg acgggggaagg 1320
tggcacggga tcgctttcct ctgaacgctt ctccgctgctc tttgagcctg cagacacctg 1380
ggggacaggg ggcgcgggat ccgctcacgg ggacagcccc cccccaaagc cccagggat 1440
gtaattacgt ccctcccccg ctagggggca cgacgagcc gcccgggggt ccgctccggg 1500
ccggcgctcc ccccgcatcc ccgagccggc agcgtgcggg gacagcccgg gcacggggaa 1560
gggggacagg gatcgctttc ctctgaacgc ttctcggtgc tctttgagcc tgcagagaa 1620
tgggggatac gggggcggcg atccgctcac ggggacagcc cccccccaaa gccccagggg 1680
atgtaattac gtccctcccc cgctaggggg cagcagcgag ccgcccgggg ctccgctccg 1740
gtccggcgct cccccgcag ccccgagccg gcagcgtgcg gggacagccc gggcacgggg 1800
aaggtggcac gggatcgctt tcctctgaac gcttctcgct gctctttgag cctgcagaca 1860
cctgggggat acggggcggg ggatccacta gttattaata gtaatcaatt acgggggtcat 1920
tagttcatag cccatatatg gagttccgcy ttacataact tacggtaaat ggcccgccgt 1980
gctgaccgcc caacgacccc cgcccattga cgtcaataat gacgtatgtt cccatagtaa 2040
cgccaatagg gaatttccat tgacgtcaat ggggtgacta tttacggtaa actgccact 2100
tggcagtaca tcaagtgtat catatgccaa gtacgcccc tattgacgtc aatgacggta 2160
aatggccggc ctggcattat gccagtaga tgaccttatg ggactttcct acttggcagt 2220
acatctacgt attagtcac tcctcccccc cctattacca tgggtcgagg tgagccccc 2280
actctcccca tctccccccc ctcccccccc ccaattttgt atttatttat tttttaatta 2340
ttttgtgcag cgatgggggc gggggggggg gggggcgcgg ccaggcgggg cggggcgggg 2400
cgaggggcgg ggccggggcg ggcggagagg tgccggcgga gccaatcaga gccgctcgct 2460
ccgaaagtgt ccttttatgg cgaggcgggg gcggcgggcg ccctataaaa agcgaagcgc 2520
gcggcgggcg ggagtgcgtg cgttgccctc gcccctgccc ccgctccgcg ccgctccgcg 2580
ccgcccgcgc cggtctgac tgaccgcgtt actcccacag gtgagcgggc gggacggccc 2640
ttctcctccg ggctgtaatt agcgtttggg ttaatgacgg ctcgtttctt ttctgtgggt 2700
cgctgaaagc cttaaagggc tccgggaggg ccctttgtgc gggggggagc ggctcggggt 2760
gtgctgtgct gtgtgtgtgc gtggggagcg ccgcgtgcgg cccgcgctgc cggcgggctg 2820
tgagcgctgc gggcgccggc gggggctttg tgcgctccgc gtgtgcgcga ggggagcgcg 2880
gccggggcgg gtgccccggg gtgcgggggg gggcgaggg gaacaaaggc tgcgtgcggg 2940
gtgtgtgctg ggggggggta ggcgggggtg tgggcgcggc ggtcgggctg taaccccccc 3000
ctgcaccccc ctccccgagt tgctgagcac gggccggctt cgggtgcggg gctccgtgcg 3060
gggcgtggcg cgggctcgcg cgtgcggggc ggggggtggc ggcaggtggg ggtgcgggg 3120
ggggcggggc cgctcgggc cggggagggc gggggggagg ggcgcggcg ccccgagcgc 3180
ccggcggtcg tcgagggcg ggcagccgca gccattgcct tttatggtaa tcgtgcgaga 3240
gggcgcaggg acttcccttg tcccaaatct ggcggagccg aaatctggga ggcgcgcgcg 3300

-72-

cacccccctct	agcggggcgcg	ggcgaagcgg	tgcggcgccg	gcaggaagga	aatggggcggg	3360
gagggccttc	gtgcgtcgcc	gcgcgcgcgt	ccccttctcc	atctccagcc	tcggggctgc	3420
cgagggggga	cggctgcctt	cggggggggac	ggggcagggc	gggggttcggc	ttctggcggtg	3480
tgaccggcg	ctctagagcc	tctgctaacc	atgttcatgc	cttcttcttt	ttcctacagc	3540
tccctgggcaa	cgtgctgggt	gttgtgctgt	ctcatcattt	tggcaaagaa	ttcgccacca	3600
tggtgagcaa	ggcgaggag	ctgttcaccg	gggtggtgcc	catcctgggtc	gagctggacg	3660
gcgacgtaaa	cggccacaag	ttcagcgtgt	ccggcgaggg	cgagggcgat	gccacctacg	3720
gcaagctgac	cctgaagttc	atctgcacca	ccggcaagct	gcccgtgccc	tggcccaccc	3780
tcgtgacctac	cttgacctac	ggcgtgcagt	gcttcagccg	ctaccccgac	cacatgaagc	3840
agcacgactt	cttcaagtcc	gccatgcccc	aaggctacgt	ccaggagcgc	accatcttct	3900
tcaaggacga	cggcaactac	aagacccgcg	ccgaggtgaa	gttcgagggc	gacacctggt	3960
accactacca	cgagctgaag	ggcatcgact	tcaaggagga	cggcaacatc	ctggggcaca	4020
agctggagta	caactacaac	agccacaacg	tctatatcat	ggccgacaag	cagaagaacg	4080
gcatcaaggt	gaacttcaag	atccgccaca	acatcgagga	cggcagcgtg	cagctcgccg	4140
acactacca	gcagaacacc	cccacccggt	gctgctgccc	gctgctgccc	gacaaccact	4200
acctgagcac	ccagtccgcc	ctgagcaaag	accccaacga	gaagcgcgat	cacatggtcc	4260
tgtctggagt	cgtgaccgcc	gccgggatca	ctctcggcat	ggacgagctg	tacaagtaag	4320
aatctcactcc	tcaggctgcag	gctgcctatc	agaagtggtg	ggctggtgtg	gccaatgccc	4380
tggtcacaa	ataccactga	gatctttttc	cctctgccaa	aaattatggg	gacatcatga	4440
agcccttga	gcatctgact	tctggctaata	aaaggaaatt	tatttttcatt	gcaatagtgt	4500
gttggaaattt	tttgtgtctc	tcactcgaa	ggacatatgg	gagggcaaat	catttaaac	4560
atcagaatga	gtatttggtt	tagagtgttg	caacatatgc	catatgctgg	ctgccatgaa	4620
caaaggtggc	tataaagagg	tcacagtat	atgaaacagc	cccctgctgt	ccattcctta	4680
ttccatagaa	aagccttgac	ttgaggttag	atttttttta	tattttgttt	tgtgttattt	4740
ttttcttttaa	catccctaaa	attttcctta	catgttttac	tagccagatt	tttctctctc	4800
tcctgactac	tcccagtcac	agctgtccct	cttctcttat	gaagatccct	cgacctgcag	4860
cccaagcttg	catcgactct	ggtcgactct	agtggtatcc	ccgccccgta	tccccagggt	4920
gtctgcaggc	tcaaagagca	gcgagaagcg	ttcagaggaa	agcgatcccc	tgccaccttc	4980
cccgctgccc	gctgtgcccc	gcacgctgcc	ggctcgggga	tgccgggggga	gcgccggacc	5040
ggagcgagc	cccggcgggc	tcgctgctgc	ccctagcgg	gggagggacg	taattacatc	5100
ctggggggct	ttgggggggg	gctgtcccc	tgagcggtac	cgccggcccc	tatccccag	5160
gtgtctgcag	gctcaaagag	cagcgagaag	cgttcagagg	aaagcgatcc	cgtgccacct	5220
tccccgtgcc	cgggctgtcc	ccgcacgtg	ccggctcggg	gatgccccgg	gagcgccgga	5280
ccggagcgga	gccccggggc	gctcgtgct	gccccctagc	gggggagggga	cgtaattaca	5340
tccccggggg	ctttgggggg	gggctgtccc	cgtgagcgga	tcccgcggccc	cgtatcccc	5400
aggtgtctgc	aggctcaaag	agcagcgaga	agcgttcaga	ggaaagcgat	cccgtgccac	5460
cttccccgtg	cccgggctgt	ccccgcacgc	tcgggctcgc	gggatgcggg	ggggagcgcc	5520
gaccgagcgc	gagccccggg	cggctcgctg	ctgcccccta	gcgggggagg	gacgtaatta	5580
catccctggg	ggctttgggg	gggggctgtc	cccgtagcgc	gatccgcggc	cccgtatccc	5640
ccaggtgtct	gcaggctcaa	agagcagcga	gaagcgttca	gaggaaaagc	atcccgatcc	5700
accttccccg	tgccccgggt	gtccccgcac	gctgcggct	cggggatgcg	gggggagcgc	5760
cggacccggg	cggagccccg	ggcggtctgc	tgtgcccccc	tagcggggga	gggacgtaat	5820
tacatccctg	ggggggttgg	ggggggggct	tccccgtgag	cggatccgcg	gccccgtatc	5880
ccccaggtgt	ctgcaggctc	aaagagcagc	gagaagcgtt	cagaggaaaag	cgatccccgt	5940
ccaccttccc	cgtgccccgg	ctgtccccgc	acgctgcccc	ctcggggatg	cgggggggagc	6000
gccggaccgg	agcggagccc	cgggcggctc	gctgctgccc	cctagcgggg	gagggacgta	6060
attacatccc	tgggggcttt	gggggggggc	tgtccccgtg	agcggatccg	cggccccgta	6120
tccccaggt	gtctgcaggc	tcaaagagca	gcgagaagcg	ttcagaggaa	agcgatcccc	6180
tgccaccttc	cccgtgcccc	ggctgtcccc	gcacgctgcc	ggctcgggga	tgccgggggga	6240
gcgccggacc	ggagcggagc	cccggggcgc	tcgctgctgc	cccctagcgg	gggagggacg	6300
taattacatc	cctgggggct	ttgggggggg	gctgtccccg	tgagcggatc	cgcggggctg	6360
caggaattcg	taatcatggt	catagctgtt	tcctgtgtga	aattgttatc	cgctcacaat	6420
tccacacaac	atagcagccg	gaagcataaa	gtgtaaagcc	tggggtgcct	aatgagttag	6480
ctaactcaca	ttaattgctg	tgcgctcact	gcccgctttc	cagtcgggaa	acctgtcgtg	6540
ccagctgcac	taatgaatcg	gccaacgcgc	ggggagaggc	gggttgctga	ttgggcgctc	6600
ttccgcttcc	tcgctcactg	actcgctgcg	ctcggtcgtt	cggctgcggc	gagcggatc	6660
agctcactca	aaggcggtaa	tacggttatc	cacagaatca	ggggataaac	caggaaagaa	6720
catgtgagca	aaaaggccagc	aaaaggccag	aaggccgcgt	tgctggcgct	tgctggcgct	6780
tttccatagg	ctccgcccc	ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagaggtg	6840
gcgaaccctg	acaggactat	aaagatacca	ggcgtttccc	cctggaagct	ccctcgtgcg	6900
ctctcctggt	cagaccctgc	cgcttaccgg	atacctgtcc	gcctttctcc	cttcgggaag	6960
cgtggcgctt	tctcatagct	cacgctgtag	gtatctcagt	tcggtgtagg	tcggtcgcctc	7020
caagctgggc	tgtgtgcacg	aacccccggt	tcagcccgac	cgctgcgcct	tatccggtaa	7080
ctatcgtctt	gagtccaacc	cggttaagaca	ccactgttcg	ccactggcag	cagccactgg	7140
taacaggatt	agcagagcga	ggtatgtagg	cggtgctaca	gagtctctga	agtgggtggc	7200
taactacggc	tacactagaa	ggacagtatt	tggatatctgc	gctctgctga	agccagttac	7260
cttcggaaaa	agagttggta	gctcttgatc	cggcaacaa	accaccgctg	gtagcgggtg	7320

-73-

tttttttgg	tgcaagcagc	agattacgcg	cagaaaaaaa	ggatctcaag	aagatccttt	7380
gatcttttct	acgggggtctg	acgctcagtg	gaacgaaaaac	tcacgttaag	ggatttttgg	7440
catgagatta	tcaaaaagga	tcttcaccta	gatcctttta	aattaaaaat	gaagttttaa	7500
atcaatctaa	agtatatatg	agtaaaacttg	gtctgcacagt	taccaatgct	taatcagtg	7560
ggcacctatc	tcagcgatct	gtctatttctg	ttcatccata	gttgccctgac	tccccgtcgt	7620
gtagataaact	acgatacggg	agggcttacc	atctggcccc	agtgtctgcaa	tgataccgcg	7680
agaccacgc	tcaccggctc	cagattttatc	agcaataaac	cagccagccg	gaagggccga	7740
gcgcagaagt	gggtcctgcaa	ctttatccgc	ctccatccag	tctattaatt	gttgccggga	7800
agctagagta	agtagttcgc	cagttaatag	tttgcgcaac	gttggttgcca	ttgctacagg	7860
catcgtgggtg	tcacgctcgt	cgtttggtat	ggcttcatte	agctccggtt	cccaacgatc	7920
aaggcgagtt	acatgatccc	ccatgttgtg	caaaaaagcg	gttagctcct	tcgggtcctcc	7980
gatcgttgct	agaagtaagt	tggccgcagt	gttatcactc	atgggttatgg	cagcactgca	8040
taattctctt	actgtcatgc	catccgtaag	atgcttttct	gtgactgggtg	agtactcaac	8100
caagtcattc	tgagaaatag	gtatgcggcg	accgagttgc	tcttgcccg	cgtcaatac	8160
ggataaatacc	gcggccacata	gcagaacttt	aaaagtgtct	atcattggaa	aacgttcttc	8220
ggggcgaaaa	ctctcaagga	tcttaccgct	gttgagatcc	agttcgatgt	aacccactcg	8280
tgaccccaac	tgatcttcag	catcttttac	tttcaccagc	gtttctgggt	gagcaaaaaac	8340
aggaaggcaa	aatgccgcaa	aaaagggaa	aaggcgaca	cggaaatgtt	gaatactcat	8400
actcttcctt	tttcaatatt	attgaagcat	ttatcagggt	tattgtctca	tgagcggata	8460
catatttgaa	tgtatttaga	aaaataaaca	aataggggtt	ccgcgcacat	ttccccgaaa	8520
agtgccacct	gacgtagtta	acaaaaaaaa	gcccgcgcaa	gcgggcttta	ttaccaagcg	8580
aagcgccatt	cgccattcag	gctgcgcaac	tgttggaag	ggcgatcggt	gcgggcctct	8640
tcgctattac	gccagctggc	gaaaggggga	tgtgctgcaa	ggcgattaa	ttgggtaacg	8700
ccagggtttt	ccagctcag	acgttgtaaa	acgacggcca	gtccgtaata	cgactcactt	8760
aaggccttga	ctagagggtc	gacgggtatc	agacatgata	agatacattg	atgagtttgg	8820
acaaaccaca	actagaatgc	agtgaaaaaa	atgctttatt	tgtgaaattt	gtgatgtctat	8880
tgctttattt	taaacctta	taagctgcaa	taaccaagtt	gggggtggcg	aagaactcca	8940
gcatgagatc	cccgcgctgg	aggatcatcc	agccggcgct	ccggaaaacg	attccgaagc	9000
ccaacctttc	atagaaggcg	gcgggtggaat	cgaaatctcg	tagcacgtgt	cagtcctgct	9060
cctcgccac	gaagtgcacg					9080

<210> 111

<211> 4223

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attBBSRpolyA10 Plasmid

<400> 111

gttaactacg	tcaggtggca	cttttcgggg	aaatgtgcgc	ggaaccctta	tttgtttatt	60
tttctaaata	cattcaaata	tgtatcgct	catgagacaa	taacctgat	aaatgcttca	120
ataatattga	aaaaggaaga	gtatagtat	tcaacatttc	cgtgtcgccc	ttattccctt	180
ttttgcggca	ttttgccttc	ctgtttttgc	tcaccagaa	acgctgggtga	aagtaaaaga	240
tgctgaagat	cagttgggtg	cacgagtggtg	ttacatcgaa	ctggatctca	acagcggtaa	300
gatccttgag	agttttcgcc	ccgaagaacg	ttctccaatg	atgagcactt	ttaaagtctt	360
gctatgtggc	gcggtattat	cccgtgttga	cgccgggcaa	gagcaactcg	gtcgcgcgat	420
acactattct	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	480
tggcatgaca	gtaagagaat	tatgcagtgc	tgccataacc	atgagtata	acactgcggc	540
caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	tgcaacaacat	600
gggggatcat	gtaactcgcc	ttgatcgttg	ggaaccggag	ctgaatgaag	ccataccaaa	660
cgacgagcgt	gacaccacga	tgccgtgtagc	aatggcaaca	acgtttgcga	aactattaac	720
tggcgaaacta	cttactctag	cttcccggca	acaattaata	gactggatgg	aggcggataa	780
agttgcagga	ccacttctgc	gctcgccctt	tccggctggc	tggtttattg	ctgataaatc	840
tggagccggt	gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	atggttaagc	900
ctcccgatc	gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	960
acagatcgct	gagatagggt	cctcactgat	taagcattgg	taactgtcag	accaagttaa	1020
ctcatatata	ctttagattg	atttaacccc	gttgataatc	agaaaaagccc	caaaaacagg	1080
aagattgtat	aagcaaatat	ttaaattgta	aacgttaata	ttttgttaaa	attcgcgtta	1140
aattttttgt	aaatcagctc	attttttaac	caataggccg	aaatcggcaa	aatcccttat	1200
aaatcaaaag	aatagccga	gatagggttg	aggtttgttc	cagtttgtaa	caagagctca	1260
ctattaaaga	acgtggactc	caacgtcaaa	gggcgaaaaa	ccgtctatca	gggcgatggc	1320
ccactacgtg	aaccatcacc	caaatcaagt	tttttggggg	cgaggtgccg	taaagcacta	1380
aatcggaacc	ctaaaggag	ccccgattt	agagcttgac	ggggaaaagcg	aacgtggcga	1440
gaaaggaagg	gaagaaagcg	aaaggagcgg	gcgctagggc	gctggcaagt	gtagcgggtca	1500
cgctgcgcgt	aaccaccaca	cccgcgcgc	ttaatgcgcc	gctacagggc	gcgtaaaagg	1560
atctaggtga	agatcctttt	tgataatctc	atgacaaaaa	tcccttaacg	tgagttttcg	1620

-74-

```

ttccactgag cgtcagaccc cgtagaaaaag atcaaaggat cttcttgaga tccttttttt 1680
ctgcgcgtaa tctgctgctt gcaaacaaaa aaaccaccgc taccagcggg gggttggttg 1740
ccggatcaag agctaccaac tctttttccg aaggtaactg gcttcagcag agcgcagata 1800
ccaaatactg ttcttctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca 1860
ccgctacat acctcgctct gctaactctg ttaccagtgg ctgctgccag tggcgataag 1920
tcgtgtctta ccgggttggg ctcaagacga tagttaccgg ataaggcgca gcggctgggc 1980
tgaacggggg gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga 2040
tacctacagc gtgagctatg agaaagcgcc acgcttcccg aaggagagaaa ggcggacagg 2100
tatccggtaa gcggcagggg cggaacagga gagcgcacga gggagcttcc aggggggaaac 2160
gcctgggtatc tttatagctc tgctcggttt cgccacctct gacttgagcg tccgtttttg 2220
tgatgctcgt cagggggggc gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg 2280
ttcctggcct tttgctggcc ttttgctcac atgtaatgtg agttagctca ctctattaggc 2340
accccaggct ttacacttta tgcttccggc tctgtatgtt tgtggaattg tgagcggata 2400
acaatttcac acaggaaca gctatgacca ggaatcgcc ttctctgcag gattgaagcc 2460
ctagtggggc ccgtgcaatt gaagccggct ggatcaccat gaaaacattt aacatttctc 2520
tgctttttta tactaacttg agcgaaatct gagagaagat tacaatgctt tatgaggata 2580
ataaacatca tgtgggagcg gcaattcgta cgaaaacagg agaatcatt tcggcagtag 2640
atattgaagc gtatatagga cgagtaactg tttgtgcaga agccattgag atttggtagt 2700
cagtttcgaa tggacaaaag gattttgaca cgattgtagc tggttagacac ccttattctg 2760
acgaagtaga cgaagtgtta gccttgggg tatgtgtagg tatgtgtagg 2820
cagactatgc accagattgt tttgtgttaa tagaaatgaa tggcaagtta gtcaaaacta 2880
cgattgaaga actcattcca ctcaaatata cccgaaatta aaagttttac cataccaagc 2940
ttggtgctg cctgaggctg gacgacctcg cggaattcta cgggagtgcc aaatccgtcg 3000
gcatccagga aaccagcagc ggctatccgc gcatccatgc ccccgactg caggagtggg 3060
gaggcacgat ggccgctttg gtccggatct ttgtgaagga accttacttc tgtgggtgtg 3120
cataatttga caaactacct acagagattt aaagctctaa ggtaaatata aaatttttaa 3180
gtgtataatg tgttaaaacta ctgattctaa ttgtttgtgt attttagatt ccaacctatg 3240
gaactgatga atgggagcag tgggtggaatg cctttaatga ggaaaacctg ttttgctcag 3300
aagaaatgcc atctagtgat gatgaggcta ctgctgactc tcaacattct actcctccaa 3360
aaaagaagag aaaggtagaa gaccccaagg actttccttc agaattgcta agttttttga 3420
gtcatgctgt gtttagtaat agaactcttg cttgctttgc tatttacacc acaaaggaaa 3480
aatagcatca gctatacaag aaaattatgg aaaaatattc tgtaaccttt ataagtaggc 3540
ataacagtta taatcataac atactgtttt ttcttactcc acacaggcat agagtgtctg 3600
ctattaataa ctatgctcaa aaattgtgtg cctttagctt ttttaattgt aaaggggtta 3660
aatagcatca caaatttcac agtgccctga ctagagatca taatcagcca taccacattt 3720
gtagagggtt tacttgcttt aaaaaacctc ccacacctcc cctgaacct gaaacataaa 3780
atgaatgcaa ttgttgttgt taacttggtt attgcagctt ataatgggtt caaataaagc 3840
aatagcatca caaatttcac aaataaagat ccacgaattc gctagcttcg gccgtgacgc 3900
gtctccggat gtacaggcat gcgtcgaccc tctagtcaag gccttaagtg agtcgtatta 3960
cggactggcc gtcgttttac aacgtcgtga ctgggaaaac cctggcggtt agtcgtatta 4020
tcgctttgca gcacatcccc ctttcgcagc ttggcgtaat cctggcggtt cccaacttaa 4080
tcgcccctcc caacagttgc gcagcctgaa agcggaaggg agcggaaggg cccgcaccga 4140
cccgttcgg cgggcttttt ttt tggcgaatgg cgcttcgctt ggtaataaag 4200

```

<210> 112
 <211> 5855
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pCX-LamIntR Plasmid

```

<400> 112
gtcgacattg attattgact agttattaat agtaatcaat tacgggggtca ttagttcata 60
gcccataatg ggagttccgc gttacataac ttacggtaaa tggcccgccg ggctgaccgc 120
ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180
ggactttcca ttgacgtcaa tgggtggact atttacggta aactgccac ttggcagtag 240
atcaagtgtg tcatatgcca agtacgcccc ctattgacgt caatgacggg aaatggcccc 300
cctggcatta tgcccagtag atgaccttat gggacttttc tacttggcag tacatctacc 360
tattagtcat cgctattacc atgggtcgag gtgagcccca cgttctgctt cactctcccc 420
atctcccccc cctccccacc cccaattttg tattttatta ttttttaatt attttgtgca 480
gggatggggg cggggggggg gggggcgccg gccaggcggg gcggggcggg gcgagggggc 540
gggcgggggc gtgcccggag agccaatcag agcggcgccg tccgaaagtt tccgaaagtt 600
tccttttatg gcgagggcgg gccctataaa aagcgaaggc cgcggcgggc cgcggcgggc 660
gggagtcgct gcgttgctt cgcctccgac cccgctccgc gccgctccgc gccgccccgc 720
ccggctctga ctgaccgcgt tactccaca ggtgagcggg cgggacggcc cttctctccc 780

```

gggctgtaat	tagcgcttgg	tttaatgacg	gctcgtttct	tttctgtggc	tgcgtgaaag	840
ccttaaaggg	ctccgggagg	gccctttgtg	cgggggggag	cggctcgggg	gggtgcgtgcg	900
tgtgtgtgtg	cgtggggagc	gccgcgtgcg	gcccgcgctg	cccgccggct	gtgagcgctg	960
cgggcgcgcc	gcggggcttt	gtgcgtcccg	cgtgtgcgcg	aggggagcgc	ggccggggggc	1020
gggtgccccg	gggtgcgggg	ggctgcgagg	ggaacaaagg	ctgcgtgcgg	gggtgtgtgcg	1080
tgggggggtg	agcagggggg	gtgggcgcgg	cggtcgggct	gtaacccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cggccccggc	tccgggtgcg	ggctccgtgc	ggggcggtggc	1200
gcggggctcg	ccgtgcgggg	cgggggggtg	cggcagggtg	gggtgccggg	cggggcgggg	1260
ccgcctcggt	ctcggggagg	ctcgggggag	gggcgcggcg	gccccggagc	gccggcggtc	1320
gtcgaggcgc	ggcgagccgc	agccattgcc	ttttatggta	atcggtgcgag	agggcgcgagg	1380
gacttccttt	gtcccaaatc	tggcgggagc	gaaactctgg	aggcgccgcc	gcacccccctc	1440
tagcggggcg	gggcgaagcg	gtgcggcgcc	ggcagggaag	aaatgggcgg	ggaggggcctt	1500
cgtgcgtgcg	cgcgcgcgcc	tccccctctc	catctccagc	ctcggggctg	ccgcagggggg	1560
acggctcgct	tccggggggg	cggggcgagg	cgggggttcg	cttctggcgt	gtgacggcg	1620
gctctagagc	ctctgctaac	catgttcatg	ccttcttctt	tttctacag	ctcctgggca	1680
acgtgctggg	tgtgtgtgctg	tctcatcatt	ttggcaaaaga	attcatggga	agaaggcgaa	1740
gtcatgagcg	ccgggattta	ccccctaacc	tttatataag	aaacaatgga	tattactgct	1800
acaggggacc	aaggacgggt	aaagagtttg	gattaggcag	agacaggcga	atcgcaatca	1860
ctgaagctat	acaggccaac	attgagttat	tttcaggaca	caaacacaag	cctctgacag	1920
cgagaatcaa	cagtgataat	tccgttacgt	tacattcatg	gcttgatcgc	tacgaaaaaa	1980
tccggggcgc	cagagggaatc	aagcagaaga	cactcataaa	ttacatgagc	aaaattaaag	2040
caataaggag	gggtctgcct	gatgctccac	ttgaagacat	caccacaaaa	gaaattgcgg	2100
caatgctcaa	tggatacata	gacgagggca	aggcggcgct	agccaagtta	atcagatcaa	2160
cactgagcga	tgcattccga	gaggcaatag	ctgaaggcca	tataacaaca	aacctgtcgc	2220
ctgccactcg	cgcagcaaaa	tctagagtaa	ggagatcaag	acttacggct	gacgaatacc	2280
tgaaaaattta	tcaagcagca	gaatcatcac	catgttgggt	cagacttgca	atggaaactgg	2340
ctgtttgttac	cggggcaacga	gttgggtgatt	tatgcgaaat	gaagtggctc	gatatcgtag	2400
atggatatct	ttatgtcgag	caaagcaaaa	caggcgtaaa	aattgoccatc	ccaacagcat	2460
tgcataattga	tgtctccgga	atatcaatga	aggaaacact	tgataaatgc	aaagagattc	2520
ttggcggaga	aaccataatt	gcactctact	ctcgcgaaacc	gctttcatcc	ggcacagtat	2580
caaggatattt	tatgcgcgca	cgaaaagcat	caggctcttc	cttcgaaggg	gatccgccta	2640
cctttcacga	gtttgcgcagt	ttgtctgcaa	gactctatga	gaagcagata	agcgataagt	2700
ttgtccaaca	tcttctcggg	cataagtcgg	acaccatggc	atcacagtat	cgtgatgaca	2760
gaggcaggga	gtgggacaaa	attgaaatca	aataagaatt	cactcctcag	gtgcaggctg	2820
cctatcagaa	gggtgggtgg	gggtgtggcca	atgccctggc	tcacaaatac	cactgagatc	2880
tttttccctc	tgcccaaaaat	tatggggaca	tcatggaacc	ccttgagcat	ctgacttctg	2940
gctaataaag	gaaatttatt	ttcattgcaa	tagtgtgttg	gaattttttg	tgtctctcac	3000
tccgaaggac	atatggggagg	gcaaatcatt	taaaacatca	gaatgagtat	ttgggtttaga	3060
gtttggcaac	atatggccata	tgctggctgc	ctcgcaaaa	gggtggctata	aagagctcat	3120
cagtatatga	aacagccccc	tgctgtccat	tccttatctc	atagaaaagc	cttgacttga	3180
ggttagatttt	tttttatatt	ttgtttttgtg	ttattttttt	ctttaacatc	cctaaaaattt	3240
tcctttacatg	ttttactagc	cagatttttt	ctcctctcct	gactactccc	agtcatagtc	3300
gtccctcttc	tcttatgaag	atccctcgag	ctgcagccca	agcttgccgt	aatcatgggtc	3360
atagctgttt	cctgtgtgaa	attgttatcc	gctcacaatt	ccacacaaca	tacgagccggg	3420
aagcataaag	tgtaaagcct	ggggtgccta	atgagtgcgc	taactcacat	taattgcgtt	3480
gcgctcactg	cccgttttcc	agtcgggaaa	cctgtcgtgc	cagcggatcc	gcactctcaat	3540
tagtcagcaa	ccatagtcct	gccccctaact	ccgcccctcc	cgcccctaac	tccgcccagt	3600
tccgcccatt	ctccgcccc	tggctgacta	atttttttta	tttatgcaga	ggccgaggcc	3660
gctcgggct	ctgagctatt	ccagaagtag	tgaggaggct	tttttggagg	cctaggcttt	3720
tgcaaaaagc	taacttgttt	attgcagctt	ataatggtta	caaataaagc	aatagcatca	3780
caaatttcac	aaataaagca	ttttttttcac	tgcatcttag	ttgtgggttg	tccaaactca	3840
tcaatgtatc	ttatcatgtc	tggatccgct	gcattaatga	atcgcccaac	gcgcgggggag	3900
aggcgggttg	cgtattgggc	gctcttccgc	ttcctcgctc	actgactcgc	tgcgctcggt	3960
cgttcggctg	cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	4020
atcaggggat	aacgcaggaa	agaacatgtg	agcaaaaagg	cagcaaaaagg	ccaggaaccg	4080
taaaaaggcc	gcgttgctgg	cgtttttcca	taggctccgc	ccccctgacg	agcatcacaa	4140
aaatcgacgc	tcaagtccga	gggtggcgaaa	cccgacagga	ctataaagat	accaggcggt	4200
tcccctcgga	agctccctcg	tgcgctctcc	tgctccgacc	ctgcgcgtta	ccggatacct	4260
gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctcaa	tgtccacgct	gtagggtatct	4320
cagttcgggtg	taggtcgttc	gctccaagct	gggtgtgtgtg	cacgaacccc	ccgttcagcc	4380
cgaccgctgc	gtaactatcg	gttaactatcg	tcttgagtc	aaccgggtaa	gacacgactt	4440
atcgccactg	gcagcagcca	ctggtaacag	gattagcaga	gcgagggtatg	taggcggtgc	4500
tacagagttc	ttgaagtggg	ggcctaacta	cggctacact	agaaggacag	tattttggtat	4560
ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	4620
acaaaccacc	gctggtagcg	gtgggttttt	tggttgcaag	cagcagatta	cgcgcagaaa	4680
aaaaggatct	caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	agtggaaacga	4740
aaactcacgt	taagggatct	tgggtcatgag	attatcaaaa	aggatcttca	cctagatcct	4800

-76-

```

tttaaattaa aaatgaagtt ttaaataaat ctaaagtata tatgagtaaa cttgggtctga 4860
cagttaccaaa tgcttaataca gtgaggcacc tatctcagcg atctgtctat ttcgttcatc 4920
catagttgcc tgactccccg tcgtgtagat aactacgata cgggaggggt taccatctgg 4980
ccccagtgct gcaatgatac cgcgagaccc acgctcaccg gctccagatt tatcagcaat 5040
aaaccagcca gccggaaggg ccgagcgcag aagtggctct gcaactttat ccgcctccat 5100
ccagtcattt aattgttgcc ggggaagctag agtaagtagt tcgccagtta atagtttgcg 5160
caacgttggt gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc 5220
attcagctcc gggtcccaac gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa 5280
agcgggttagc tccttcgggtc ctccgatcgt tgtcagaagt aagttggccg cagtgttatc 5340
actcatgggt atggcagcac tgcataattc tcttactgtc atgccatccg taagatgctt 5400
ttctgtgact ggtgagtact caaccaagtc attctgagaa tagtgtatgc ggcgaccgag 5460
ttgctcttgc ccggcggtcaa tacgggataaa taccgcgcca catagcagaa ctttaaaagt 5520
gctcatcatt ggaaaacggt cttcggggcg aaaactctca aggatcttac cgctgttgag 5580
atccagttcg atgtaaccca ctctgcacc caactgatct tcagcatctt ttactttcac 5640
cagcgtttct gggtagagcaa aaacaggaag gcaaaaaagg gaataagggg 5700
gacacggaaa tgttgaatac tcatactctt cttttttcaa tattattgaa gcatttatca 5760
gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata aacaaatagg 5820
ggttccgcgc acatttcccc gaaaagtgcc acctg 5855

```

<210> 113

<211> 4346

<212> DNA

<213> Artificial Sequence

<220>

<223> pSV40-193AttpsensePur Plasmid

<400> 113

```

ccggtgcccgc caccatcccc tgacccacgc cctgaccccc tcacaaggag acgaccttcc 60
atgaccagagt acaagccac ggtgcgcctc gccaccgcgc acgacgtccc ccgggcccgt 120
cgacccctgt ccgcccgcgt cgcgcactac cccgccacgc gccacaccgt cgacccggac 180
cgccacatcg agcgggtcac cgaagctgcaa gaactcttcc tcacgcgcgt cggggtcgac 240
atcggaagg tgtgggtcgc ggacgacggc gccgcggtgg cggctctggac cagcccgag 300
agcgtcgaa cggggggcgt gttcgcgcag atcggcccg ccatggccga gttgagcgg 360
tcccggctgg ccgcccagca acagatggaa ggccctcctg cgccgcaccg gcccaaggag 420
cccgcgtggt tcctggccac cgtcggcgct tcgcccgacc accagggcaa gggctctggc 480
agcgcgctgc tgcctcccgg agtggaggcg gccgagcgcg ccgggggtgc cgccttctct 540
gagacctccg cgccccgcaa cctccccttc tacgagcggc tcggcttcac cgtcacccgc 600
gacgtcgagg tgcccgaagg accgcgcacc tgggtgcatg cccgcaagcc cgggtgctga 660
cgcccccccc acgacccgca cgcgccgacc gaaaggagcg cagcacccca tgggtccgag 720
cgaaagccgac ccggggcgcc ccgcccagccc cgcacccgcc cccgaggccc accgactcta 780
gaggatcata atcagccata ccacatttgt agaggtttta cttgctttta aaaacctccc 840
acacctcccc ctgaacctga aacataaaat gaatgcaatt gttgttgta acttgtttat 900
tgacgttat aatgggtaca aataaagcaa tagcatcaca aatttcacaa ataaagcatt 960
tttttactgc cattctagtt gtgggtttgt caaactcatc aatgtatctt atcatgtctg 1020
gatccgcgcg ggatccttaa ttaagtctag agtcgactgt ttaaacctgc aggcattgcaa 1080
gcttgccgta atcatggtca tagctgtttc ctgtgtgaaa ttgttatccg ctcacaattc 1140
cacacaacat acgagccgga agcataaagt gtaaagcctg ggggtgcctaa tgagttagct 1200
aactcacatt aattgcgttg cgtcactgc ccgctttcca gtcgggaaac ctgtcgtgac 1260
agctgcatta atgaatcggc caacgcgcgg ggagaggcgg tttgcgtatt gggcgctctt 1320
ccgcttcctc gctcactgac tcgctgcgct cggctcgttc gctgcggcga gcggtatcag 1380
ctcactcaaa ggcggtaata cggttatcca cagaatcagg ggataacgca ggaaagaaca 1440
tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa ggccgcgttg ctggcggttt 1500
tccataggct ccgccccctt gacgagcatc acaaaaaatc acgctcaagt cagaggtggc 1560
gaaacccgac aggaactata agataccagg cgtttccccc tggaaagctc ctcgtcgct 1620
ctcctgttcc gacctgccc cttaccggat acctgtccgc cttctccctc tcgggaagcg 1680
tggcgctttc tcatagctca cgctgtaggt atctcagttc ggtgtaggtc gttcgctcca 1740
agctgggctg tgtgcacgaa cccccgcttc agtccgaccg ctgcgcctta tccggtaact 1800
atcgtcttga gtccaacccg gtaagacacg acttatcgcc actggcagca gccactggta 1860
acaggattag cagagcgagg tatgtaggcg gtgctacaga gttcttgaag tgggtggccta 1920
actacggcta cactagaagg acagtatttg gtatctgcgc tctgctgaag ccagttacct 1980
tcggaaaaag agttggtagc tcttgatccg gcaaaacaac caccgctggg agcgggtgggt 2040
tttttgtttg caagcagcag attacgcgca gaaaaaaaag atctcaagaa gatcctttga 2100
cttttctac ggggtctgac gctcagtgga acgaaaaactc acgttaaggg attttggta 2160
tgagattatc aaaaaggatc ttcacctaga tctttttaa ttaaaaatga agttttaa 2220
caatctaaag tatatatgag taaacttggt ctgacagtta ccaatgctta atcagttagg 2280
cacctatctc agcgatctgt ctatttcggt catccatagt tgcctgactc ccgctcgtgt 2340

```


-77-

```

agataactac gatacggggag ggcttaccat ctggccccag tgctgcaatg ataccgcgag 2400
acccacgctc accgggtcca gatttatcag caataaacca gccagccgga agggccgagc 2460
gcagaagtgg tccgtgcaact ttatccgcct ccatccagtc tattaattgt tgccgggaag 2520
ctagagtaag tagttcgcca gttaatagtt tgccgaacgt tggtgccatt gctacaggca 2580
tcgtgggtgc acgctcgctc tttgggtatgg ctccattcag ctccggttcc caacgatcaa 2640
ggcgagttac atgatcccc atgttgtgca aaaaagcggg tagctccttc ggtcctccga 2700
tcgttgtcag aagtaagttg gccgcagtggt tatcactcat ggttatggca gactgcata 2760
attctcttac tgtcatgcca tccgtaagat gcttttctgt gactggtag gactcaacca 2820
agtcattctg agaatagtggt atgcggcgac cgagttgctc ttgcccggcg tcaatacggg 2880
ataataccgc gccacatagc agaactttaa aagtgtctcat cattggaaaa cgttcttcgg 2940
ggcgaaaaact ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa cccactcggt 3000
caccacaactg atcttcagca tcttttactt tcaccagcgt ttctgggtga gcaaaaaacag 3060
gaaggcaaaa tgccgcacaaa aagggaataa gggcgacacg gaaatgttga atactcatac 3120
tcttcctttt tcaatattat tgaagcattt atcagggtta ttgtctcatg agcggataca 3180
tatttgaaatg tatttagaaa aataaacaaa taggggttcc gcgcacattt ccccgaaaag 3240
tgccacctga cgtctaagaa accattatta tcatgacatt aacctataaa aataggcgta 3300
tcacgagggc ctcttcgtctc gcgcgtttcg gtgatgacgg tgaaaacctc tgacacatgc 3360
agctcccggg gacgggtcaca gcttgtctgt aagcggatgc cgggagcaga caagcccgtc 3420
agggcgcgctc agcgggtgtt ggccgggtgtc ggggctggct taactatgct gcacagatgc 3480
agattgtact gagagtgcac catatgcggt gtgaaatacc gcacagatgc gtaaggagaa 3540
aatcccgcat caggcgccat tccgccattc gctgcgcaa ctgttgggaa gggcgatcgg 3600
tgccggcctc ttccgtatta cgccagctgg cgaaaggggg atgtgtgca aggcgattaa 3660
gttgggtaac gccagggttt tcccagtcac gacgttgtaa aacgacggcc agtgaattcg 3720
agctgtggaa tgtgtgtcag ttagggtgtg gaaagtcccc aggcctccca gaggcgagaa 3780
gtatgcaaaag catgcatctc aattagtcag caaccagggtg tggaaagtcc ccaggctccc 3840
cagcagggcag aagtatgcaa agcatgcatt ccaattagtc agcaaccata gtcccggccc 3900
taactccgct ctaaccgccc gcagagggcc aggttccgc ccatctccg cccatggct 3960
gactaatttt ttttatttat gcagagggcc agggccgctc ggctctctgag ctattccaga 4020
agtagtgagg aggccttttt ggaggctcgg taccoccttg cgctaattgct ctgttacagg 4080
tcactaatac catctaatga gttgattcat agtgactgca tatgttgtgt tttacagtat 4140
tatgtagctc gttttttatg caaaatctaa tttacatat tgatatttat atcattttac 4200
gtttctcggt cagctttttt atactaagtt ggcattataa aaaagcattg cttatcaatt 4260
tgttgcaacg aacaggctac tatcagtc aaataaatca ttatttgatt tcaattttgt 4320
ccactccct gcctctgggg ggcgcg 4346

```

<210> 114

<211> 3166

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeo Plasmid

<400> 114

```

cagttgccgg ccgggtcgcg cagggcgaaac tccgcggccc acggctgctc gccgatctcg 60
gtcatggccg gcccggaggc gtcccggaaag ttctgtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggcccgcg caccacacacc caggccaggg tggtgtccgg caccacctgg 180
tccgtggaccg cgctgatgaa caggggtcacg tcgtcccggg ccacaccggc gaagtctcc 240
tccacgaagt cccgggagaa cccgagccgg tccgtccaga actcgaccgc tccggcgacg 300
tcgcgcgcgg tgagcaccgg aacggcactg gtcaacttgg ccattggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgcattg ctgcaggctc 420
actctagagg atccccgggt accgagctcg aattcgtaat catggtcata gctgtttcct 480
gtgtgaaatt gttatccgct cacaattcca cacaacatac gagccggaag cataaagtgt 540
aaagcctggg gtgcctaatt agtgagctaa ctacattaa ttgcgttgcg ctactgccc 600
gctttccagt cgggaaacct gtctgtccag ctgcattaat gaatcggeca acgcgcgggg 660
agaggcgggt tgctgattgg gcgctcttcc gcttccctgc tcaactgactc gctgcgctcg 720
gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg cggtaatacg gttatccaca 780
gaatcagggg ataacgcagg aaagaacatg tgagcaaaa gccagcaaaa ggccaggaa 840
cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc gccccctga cgagcatcac 900
aaaaatcgac gctcaagtcg gaggtggcga aaccgcagag gactataaag ataccaggcg 960
tttcccctcg gaagctccct cgtgcgctct cctgttccga ccctgcccgt taccggatac 1020
ctgtccgctt ttctcccttc gggaagcggt gcgctttctc atagctcacg ctgtagggtat 1080
ctcagttcgg tgtaggctcg tcgctccaag ctgggctgtg tgcacgaacc cccggttcag 1140
ccccagcgtc cggtacttat cggtaactat cgtcttgagt ccaacccggg aagcacgacg 1200
ttatcgccac tggcagcagc cactggtaac aggtatagca gagcgaggta tgtaggcggt 1260
gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaaggac agtatttggt 1320
atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc 1380

```

-78-

aaacaaacca	ccgctggtag	cgggtggtttt	tttgtttgca	agcagcagat	tacgcgcaga	1440
aaaaaaggat	ctcaagaaga	tcctttgatc	ttttctacgg	ggtctgacgc	tcagtggaaac	1500
gaaaactcac	gttaagggat	tttgggtcatg	agattatcaa	aaaggatctt	cacctagatc	1560
cttttaaat	aaaaatgaag	ttttaaatca	atctaaagta	tatatgagta	aacttgggtct	1620
gacagttacc	aatgcttaat	cagtgaaggca	cctatctcag	cgatctgtct	atttcgttca	1680
tccatagtgtg	cctgactccc	cgctgtgtag	ataactacga	tacgggaggg	cttaccatct	1740
ggccccagtg	ctgcaatgat	accgcgagac	ccacgctcac	cggtcccaga	tttatcagca	1800
ataaaccagc	cagccggaag	ggccgagcgc	agaagtgggtc	ctgcaacttt	atccgcctcc	1860
atccaggtcta	ttaattgttg	ccgggaagct	agagtaagta	gttcgcccagt	taatagtgtt	1920
cgcaacgttg	ttgccattgc	tacaggcatc	gtgggtgtcac	gctcgtcgtt	tggtatggct	1980
tcattcagct	ccgggttccca	acgatcaagg	cgagttacat	gatcccccat	gttgtgcaaa	2040
aaagcgggta	gctccttcgg	tcctccgatc	gttgtcagaa	gtaagtgggc	cgagtggtta	2100
tcactcatgg	ttatggcagc	actgcataat	tctcttactg	tcatgccatc	cgtaagatgc	2160
ttttctgtga	ctgggtgagta	ctcaaccaag	tcattctgag	aatagtgtat	gcggcgaccg	2220
agtgtctctt	gcccggcgtc	aataccgcgc	gccaacgcgc	cacatagcag	aactttaaaa	2280
gtgctcatca	ttggaaaacg	ttcttcgggg	cgaaaactct	caaggatctt	accgctgttg	2340
agatccagtt	cgatgtaacc	cactcgtgca	cccaactgat	cttcagcatc	ttttactttt	2400
accagcgttt	ctgggtgagc	aaaaacagga	aggcaaaatg	ccgcaaaaaa	gggaatagg	2460
gcgacacgga	aatggtgaat	actcatactc	ttcctttttc	aatattattg	aagcatttat	2520
cagggttatt	gtctcatgag	cggatacata	tttgaatgta	tttagaaaaa	taaacaaata	2580
ggggttcgcg	gcacatttcc	ccgaaaagtgc	ccacctgacg	tagttaacaa	aaaaaagccc	2640
gccgaagcgg	gctttattac	caagcgaagc	gccattcgcc	attcaggctg	cgcaactggt	2700
gggaagggcg	atcggtgcgg	gcctcttcgc	tattacgcca	gctggcgaaa	gggggatgtg	2760
ctgcaaggcg	attaagttag	gtaacgccag	ggttttccca	gtcacgacgt	tgtaaaacga	2820
cgggcagttc	gtaatacgac	tcacttaagg	ccttgactag	agggtcgacg	gtatacagac	2880
atgataagat	acattgatga	gtttggacaa	accacaacta	gaatgcagtg	aaaaaaatgc	2940
tttatttgtg	aaatttgtga	tgctattgct	ttatttgtaa	ccattataag	ctgcaataaa	3000
caagttgggg	tgggcgaaga	actccagcat	gagatccccg	cgctggagga	tcattccagcc	3060
ggcgtcccgg	aaaacgattc	cgaagcccaa	cctttcatag	aaggcggcgg	tggaatcgaa	3120
atctcgtagc	acgtgtcagt	cctgctcttc	ggccacgaag	tgcacg		3166

<210> 115

<211> 7600

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeo3' 6XHS4eGFP Plasmid

<400> 115

cagttgcccg	ccgggtcgcg	cagggcggaac	tcccgcctcc	acggctgctc	gccgatctcg	60
gtcatggccg	gcccggaggc	gtcccgaag	ttcgtggaca	cgacctccga	ccactcggcg	120
tacagctcgt	ccaggccgcg	cacccacacc	caggccaggg	tggtgtccgg	caccacctgg	180
tcctggaccg	cgctgatgaa	caggggtcacg	tcgtcccggg	ccacaccggc	gaagtgcgtc	240
tcacgaagtg	cccgggagaa	cccgaagccg	tcggtccaga	actcgaccgc	tccggcgacg	300
tcgcgcgcgg	tgagcaccgg	aacggcactg	gtcaacttgg	ccatggatcc	agatttcgct	360
caagttagta	taaaaaagca	ggcttcaatc	ctgcagagaa	gcttgatcta	gttattaata	420
gtaatcaatt	acggggtcat	tagttcatag	cccataatg	gagttccgcg	ttacataact	480
tacggtaaat	ggcccgcctg	gctgaccgcc	caacgacccc	cgcccattga	cgtcaataat	540
gacgtatggt	cccatagtaa	cgccaatagg	gactttccat	tgacgtcaat	gggtggacta	600
tttacggtaa	actgcccact	tggcagtaca	tcaagtgtat	catatgccaa	gtacgcccc	660
tattgacgtc	aatgacggta	aatggcccgc	ctggcattat	gcccagtaga	tgaccttatg	720
ggactttcct	acttggcagt	acatctacgt	attagtcatt	gctattacca	tgggtcgagg	780
tgagccccac	gttctgcttc	actctcccca	tctccccccc	ctccccaccc	ccaattttgt	840
atttatttat	tttttaatta	ttttgtgcag	cgatgggggc	gggggggggg	ggggcgcgcg	900
ccaggcgggg	cggggcgggg	cgaggggcgg	ggcggggcga	ggcggagagg	tgccgcggca	960
gccaatcaga	gcggcgcgct	ccgaaagtgt	ccttttatgg	cgaggcggcg	gcggcgggcg	1020
ccctataaaa	agcgcaagcg	gcggcgggcg	ggagtcgctg	cgttgccttc	gccccgtgcc	1080
ccgctccgcg	ccgcctcgcg	ccgcccgcgc	cggtctgac	tgaccgcgtt	actcccacag	1140
gtgagcgggc	gggacggccc	ttctcctcgc	ggctgtgaat	agcgcttggt	ttaatgacgg	1200
ctcgtttctt	ttctgtggct	gcgtgaaagc	cttaaaaggg	tccgggaggg	ccctttgtgc	1260
ggggggggagc	ggctcggggg	gtgctgctgc	gtgtgtgtgc	gtggggagcg	ccgcgtgcgg	1320
cccgcgctgc	ccggcggtgc	tgagcgctgc	gggcgcggcg	cggggctttg	tgcgctccgc	1380
gtgtgcgcga	ggggggggcg	ggcgggggcg	gtgccccgcg	gtgccccggg	gctgcgaggg	1440
gaacaaaggc	tgcgtgcggg	gtgtgtgcgt	gggggggtga	gcaggggggtg	tgggcgcggc	1500
ggtcgggctg	taaccccccc	ctgcaccccc	ctccccgagt	tgctgagcac	ggccccggtt	1560
cgggtgcggg	gctccgtgcg	gggcgtggcg	cggggctcgc	cgtgcggggc	gggggggtgg	1620

ggcaggtggg ggtgcccgggc gggggcggggc cgcctcggggc cggggaggggc tcggggggagg 1680
ggcgccggcg ccccgagcg cggcgccgtg tcgagggcg cgcagccgca gccattgcct 1740
tttatggtaa tcgtgcgaga gggcgccaggg acttcctttg tcccaaactt ggcggagccg 1800
aaatctggga ggcgcgcggc caccocctct agcggggcg ggcgaagcgg tcggcgcccg 1860
gcaggaagga aatgggcggg gagggccttc gtgcgtcgcc gcgccgcgt cccctctcc 1920
atctccagcc tcggggctgc cgcaggggga cggctgcctt cggggggggac ggggcagggc 1980
gggggttcggc ttctggcggtg tgaccggcg ctctagagcc tctgctaacc atgttcatgc 2040
cttcttcttt ttctacagc tccctgggcaa cgtgctgggt gtgtgtgtgt ctcatcatt 2100
tggcaagaa ttgcaccca tggtagcaa gggcgaggag ctgttcaccg ggggtggtgcc 2160
catcctgggt gagctggacg ggcagctaaa cggccacaag ttcagcgtgt ccggcgaggg 2220
cgaggggcgt gccacctacg gcaagctgac cctgaagttc atctgcacca ccggcaagct 2280
gcccgtgccc tggcccaccc tcgtgaccac cctgacctac ggcgtgcagt gcttcaggcg 2340
ctaccccgac cacatgaagc agcacgactt cttcaagtcc gccatgcccg aaggtacgt 2400
ccaggagcgc accatcttct tcaaggacga cggcaactac aagaccgcg ccgaggtgaa 2460
gttcgagggc gacacctgg tgaaccgcat cgagctgaag ggcacgcact tcaaggagga 2520
cggcaacatc ctggggcaca agctggagta caactacaac agccacaacg tctatatcat 2580
ggccgacaag cagaagaacg gcatcaaggt gaactcaag atccgccaca acatcaggga 2640
cggcagcgtg cagctcgccg accactacca gcagaacacc cccatcgggc acggcccggt 2700
gctgctgccc gacaaccact acctgagcac ccagtcggcc ctgagcaaaag accccaacga 2760
gaagcgcgtg cacatgggtcc tgctggagtt cgtgaccgcc gccgggatca ctctcggtat 2820
ggacgagctg tacaagtaag aattcactcc tcagggtcag gctgcctatc agaaggtgg 2880
ggctgggtgt gccaatgccc tggctcaca ataccactga gatctttttc cctctgccaa 2940
aaattatggg gacatcatga agcccttga gcatctgact tctggctaata aaaggaaatt 3000
tattttcatt gcaatagtgt gttggaattt ttgtgtctc tcactcgga ggcacatagg 3060
gagggcaaat catttaaac atcagaatga gtatttgggt tagagtttgg caacatatgc 3120
catatgctgg ctgccatgaa caaagggtggc tataaagagg tcatcagtat atgaaacagc 3180
ccctgctgt ccattcctta ttccatagaa aagccttgac ttgaggttag attttttta 3240
tattttgttt tgtgttattt ttttctttaa catccctaaa attttctta catgttttac 3300
tagccagatt ttctctctc tccgtactac tccagtcact agctgtccct ctctcttat 3360
gaagatccct cgacctgcag cccaagcttg atgcctgca ggtcgactct agtggatccc 3420
ccgccccgta tccccaggt gtctgcaggc tcaaagagca ggcagaagcg ttcagaggaa 3480
agcgatcccg tgccaccttc cccgtgcccg ggcgtgcccc gcacgctgcc ggctcgggga 3540
tgccggggga gcgcgggacc ggagcggagc cccggggggc tcgctgctgc cccctagcg 3600
gggagggagc taattacatc cctggggggt ttgggggggg gctgtccccg tgagcggatc 3660
cgccggcccc tatccccag gtgtctgcag gctcaaagag cagcgagaag cgttcagagg 3720
aaagcgatcc cgtgcacct tcccgtgccc cggcctgtcc ccgcacgctg ccggtcggg 3780
gatgcggggg gagcgccgga ccggagcggg gccccggggc gctcgctgct gccccctagc 3840
ggggggggga cgtaattaca tccctggggg ctttgggggg ggggtgtccc cgtgagcgga 3900
tccgcgccc cgatcccc cgtgtctct aggtcctaa agcagcgaga agcgttcaga 3960
ggaaagcgat cccgtgccac cttccccgtg cccgggtgt ccccgacgc tgccggctcg 4020
gggatgcggg gggagcgccg gaccggagcg gagccccggg cggctcgctg ctgcccccta 4080
gcggggggag gacgtaatta catccctggg ggctttgggg gggggctgtc cccgtgagc 4140
gatccggggc cccgtatccc ccaggtgtc cagagctcaa agagcagcga gaagcgttca 4200
gaggaaagcg atcccggtgc accttcccc cggccggggt gtccccgcac gctgcccgtt 4260
cggggatgcg gggggagcgc cggaccggag cggagccccg ggcggctcgc tgtgcccc 4320
tagcggggga gggacgtaat tacatccctg gggggtttgg gggggggctg tccccgtgag 4380
cggatcccg cgccccgtat ccccaggtgt ctgcaggctc aaagagcagc gagaagcgtt 4440
cagaggaaa gcatcccggt ccaccttccc cgtgccccgg ctgtccccgc acgtgccc 4500
ctcggggatg cgggggggag gccggagccc agcggagccc cgggcgggtc gctgtgccc 4560
cctagcgggg gagggacgta attacatccc tggggggttt gggggggggc tgtccccgtg 4620
agcggatccg cgcccccgta tccccaggt gtctgcaggc tcaaagagca gcgagaagcg 4680
ttcagaggaa agcgatccc tgccaccttc cccgtgccc ggctgtcccc gcacgctgcc 4740
ggctcgggga tgccggggga ggcgcggacc ggagcggagc cccggggggc tcgctgctgc 4800
cccctagcgg gggagggagc taattacatc cctggggggt ttgggggggg gctgtccccg 4860
tgagcggatc cgcggggctg caggaattcg taatcatgg catagctgt tctgtgtga 4920
aattgttatc cgctcacaat tccacacaac atacgagccg gaagcataaa gtgtaaagcc 4980
tgggggtgct aatgagtgag ctaactcaca ttaattgcgt tgcgtcact gcccgtttc 5040
cagtccggaa acctgtcggt ccagctgcat taatgaactg gccaacgcgc ggggagaggc 5100
ggtttgcgta ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctccggtcgtt 5160
cggctgcggc gagcgggtat agctcactca aaggcggtaa tacggttatc cacagaatca 5220
ggggataacg caggaaaagaa catgtgagca aaaggccagc aaaggccagc gaaccgtaaa 5280
aaggccgcgt tgctggcggt tttccatagg ctccgcccc ctgacgagca tcacaaaaat 5340
cgacgctcaa gtcagaggtg gcgaaacccc acaggactat aaagatacca ggcgtttccc 5400
cctggaagct cctcgtgct ctctcgtgt cctaccgg cgcttaccgg atacctcagt 5460
gcctttctcc cttcggaag cgtggcgctt cctcatagct caccgtgtag gtatctcagt 5520
tcggtgtagg tcgttcgctc caagctgggc tgtgtgcacg aaccccccg tccagcccagc 5580
cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg 5640

-80-

```
ccactggcag cagccactgg taacaggatt agcagagcga ggtatgtagg cgggtgctaca 5700
gagttcttga agtgggtggc taactacggc tacactagaa ggacagtatt tggatatctgc 5760
gctctgctga agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa 5820
accacccgtg gtagcgggtgg tttttttgtt tgcaagcagc agattacgcg cagaaaaaaa 5880
ggatctcaag aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac 5940
tcacgttaag ggatttttgt catgagatta tcaaaaagga tcttcacctt gatcctttta 6000
aattaaaaat gaagttttta atcaatctaa agtatatatg agtaaaacttg gtctgacagt 6060
taccaatgct taatcagtgga ggcacctatc tcagcgatct gtctatttcg ttcacacata 6120
gttgcttgac tccccgtcgt agacccacgc tcacgggctc aggttttacc atctggcccc 6180
agtgtctgca tgataccgcg gcgcagaagt ggtcctgcaa cagatttatc agcaataaac 6240
cagccagccg gaagggccga agctagagta agtagttcgc ctttatccgc ctccatccag 6300
tctattaatt gttgccggga catcggtgtg tcacgctcgt cagttaatag tttgccgaac 6360
gttggttgcca ttgctacagg aaggcgagtt acatgatccc cgtttggtat ggcttcattc 6420
agctccggtt cccaacgatc gatcggtgtc agaagtaagt tggccgcagt gttatcactc 6480
gttagctcct tcggtcctcc taattctctt actgtcatgc catccgtaag atgcttttct 6540
atgggttatgg cagcactgca caagtcattc tgagaatagt gcagaacttt aaaagtgtc 6600
gtgactgggt agtactcaac ggataatacc ggcacacata tcttaccgct gttgagatcc 6660
tcttgcccgg cgtcaatacg ggggcgaaaa tgcattttac tttcaccagc 6720
atcattggaa aacgtttctt aggaaggcaa actcttctct tttcaatatt attgaagcat 6780
agttcgatgt aacccactcg actccttccc tttcgaatatt tttcgaatatt aataggggtt 6840
gtttctgggt gagcaaaaaa agtgccacct gacgtagtta acaaaaaaaa gcccgccgaa 6900
cggaaatgtt gaatactcat actccttccc tttcgaatatt tttcgaatatt aataggggtt 6960
tattgtctca tgagcgggata catattttgaa tgtattttaga aaaataaaca aataggggtt 7020
ccgcgcacat ttccccgaaa agtgccacct gacgtagtta acaaaaaaaa gcccgccgaa 7080
gccccgttta ttaccaagcg aagcgccatt cgccattcag gctgcgcaac tgttgggaag 7140
ggcgatcggt gccccgctct tcgctattac gccagctggc gaaaggggga tgtgctgcaa 7200
ggcgattaa gtgggtaacg ccagggtttt cccagtcacg acgttgtaaa acgacggcca 7260
gtccgtaata cgactcactt aaggccttga ctagagggtc gacgggtatac agacatgata 7320
agatacattg atgagtttgg acaaaccaca actagaatgc agtgaaaaaa atgcttttatt 7380
tgtgaaatatt gtgatgctat tgcttttatt gtaaccatta taagctgcaa taaacaagtt 7440
gggggtggcg aagaactcca gcatgagatc cccgcgctgg aggatcatcc agccggcgctc 7500
ccggaaaacg attccgaagc ccaacctttc atagaaggcg gcggtggaat cgaaatctcg 7560
tagcacgtgt cagtcctgct cctcggccac gaagtgcacg 7600
```

<210> 116

<211> 7631

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeo5' 6XHS4eGFP Plasmid

<400> 116

```
cagttgcccg ccgggtcgcg cagggcgaac tccccgcccc acgggtgctc gccgatctcg 60
gtcatggccg gcccgaggcg gtcccggaa ttcgtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggccgcg caccacacac caggccaggg tgttgtccgg caccacctgg 180
tcctggaccg cgctgatgaa cagggtcacg tcgtcccgga ccacaccggc gaagtcgtcc 240
tccacgaagt cccgggagaa cccgagcccg tcggtccaga actcgaccgc tccggcgacg 300
tcgcgccggg tgagcaccgg aacggcactg gtcaacttgg ccattggatc agatttcgct 360
caagttagta taaaaagca ggcttcaatc ctgcagagaa gcttgatatac gaattcctgc 420
agcccccgcg atccgctcac ggggacagcc cccccccaaa gcccccaggg atgtaattac 480
gtccctcccc cgtagggggg cagcagcgag ccgcccgggg ctccgctccg gtcggcgctg 540
cccccccgcat ccccgagccg gcagcgtcgc gggacagccc gggcacgggg aagggtggc 600
gggatcgctt tcctctgaac gcttctcgct gctctttgag cctgcagaca cctgggggat 660
acggggcgcg ggatccgctc acggggacag ccccccccca aagccccag ggatgtaatt 720
acgtccctcc cccgctaggg ggcagcagcg agccgcccgg ggctccgctc cgggtccggcg 780
ctcccccccg atccccgagc cggcagcgtg cggggacagc ccgggcacgg ggaaggtggc 840
acgggatcgc ttctcttga acgcttcttg ctgctctttg agcctgcaga cactgggggg 900
atacggggcc gcggatccgc tcacggggac agcccccccc caaagccccc agggatgtaa 960
ttacgtccct cccccgctag ggggcagcag cgagccgccc ggggctccgc tccgggtccg 1020
cgctcccccc gctccccga gccggcagcg cggggggaca gccggggcac ggggaagggt 1080
gcacgggatc gctttcctct gaacgcttct cgctgctctt tgagcctgca gacacctggg 1140
ggatacgggg ccggcgatcc gctcacgggg acagcccccc cccaaagccc ccagggatgt 1200
aattacgtcc ctcccccgct agggggcagc agcagaccgc cgggggctcc gctccgggtc 1260
ggcgctcccc ccgcattccc gagccggcag cgtgcgggga cagcccgggc acgggggaag 1320
tggcacggga tcgcttttct ctgaacgctt ctcgctgctc tttgagcctg cagacacctg 1380
ggggatacgg ggcccgcggt cgcgtcacgg ggacagcccc cccccaaagc ccccgaggat 1440
```

gtaattacgt	ccctcccccg	ctaggggggca	gcagcgagcc	gccccggggct	ccgctccggt	1500
ccggcgctcc	ccccgcaccc	ccgagccggc	agcgtgcggg	gacagcccg	gcacggggaa	1560
ggtgggacgg	gatcgctttc	ctctgaacgc	ttctcgctgc	tctttgagcc	tgacagacc	1620
tgggggatac	ggggcccgcg	atccgctcac	ggggacagcc	ccccccaaa	gccccaggg	1680
atgtaattac	gtccctcccc	cgctaggggg	cagcagcgag	ccgcccgggg	ctccgctccg	1740
gtccggcgct	ccccccgcac	ccccgagccg	gcagcgtgcg	gggacagccc	gggcacgggg	1800
aaggtggcac	gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	1860
cctgggggat	acggggcggg	ggatccacta	gttattaata	gtaatcaatt	acgggggtcat	1920
tagttcatag	cccatatatg	gagttccgcg	ttacataact	tacggtaaat	ggccgcgctg	1980
gctgaccgcc	caacgacccc	cgcccattga	cgtcaataat	gacgtatgtt	cccatagtaa	2040
cgccaatagg	gactttccat	tgacgtcaat	gggtggacta	tttacggtaa	actgcccact	2100
tggcagtaca	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggcccg	ctggcattat	gcccagtaga	tgaccttatg	ggactttcct	acttggcagt	2220
acatctacgt	attagtcac	gctattacca	tgagtcgagg	tgagccccac	gttctgcttc	2280
actctcccca	ctcccccccc	ctcccccccc	ccaattttgt	atttatttat	tttttaatta	2340
ttttgtgcag	cgatgggggg	gggggggggg	ggggcgcgcg	ccaggcgggg	cgggcggggg	2400
cgaggggcg	ggcggggcga	ggcggagagg	tgccggcgga	gccaatcaga	gcggcgcgct	2460
ccgaaagtgt	cccttttatg	cgaggcggg	gcggcgggcg	ccctataaaa	agcgaagcgc	2520
gcggcgggcg	ggagtcgctg	cgttgccttc	gccccgtgcc	ccgctccgcg	ccgctccgcg	2580
ccgccccccc	cggtctcgac	tgaccgcggt	actcccacag	gtgagcgggc	gggacggccc	2640
ttctcctccg	ggctgtaat	agcgtttggt	ttaatgacgg	ctcgtttctt	ttctgtcggt	2700
gcgtgaaagc	cttaaagggc	tcggggaggg	ccctttgtgc	gggggggagc	ggctcggggg	2760
gtgcgtgcgt	gtgtgtgtgc	gtggggagcg	ccgcgtgcgg	ccgcgctgc	ccggcgggctg	2820
tgagcgctgc	gggcgcggcg	cggggctttg	tcgcgtccgc	gtgtgcgcga	ggggagcgcg	2880
gccggggggcg	gtgccccgcg	gtgcgggggg	gctgcgaggg	gaacaaaggc	tgctgcgggg	2940
gtgtgtgcgt	gggggggtga	gcaggggggt	tgggcgcggg	ggtcgggctg	taaccccccc	3000
ctgcaccccc	ctccccgagt	tgctgagcac	ggccccgctt	cggtgcgggg	gctccgtgcg	3060
gggctgtggc	cggggctcgc	cggtgccggg	ggggggtggc	ggcaggtggg	ggtgccgggg	3120
ggggcggggg	cgcctcgggg	cggggagggc	tcgggggagg	ggcgcgcgcg	ccccggagcg	3180
ccggcgctgc	tcagggcgcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcggag	3240
ggggcgaggg	acttctcttg	tcaccaaatc	ggcgagcccg	aaatctggga	ggcgccggcg	3300
cacccccctc	agcggggcg	ggcgaagcgg	tgccggcgcg	gcaggaaagg	aatggggcg	3360
gaggggcttc	gtgcgtcgcc	gcgcgcggcg	ccccttctcc	atctccagcc	tcggagcgcg	3420
cgcaggggga	cggtctcctt	cgggggggac	ggggcagggc	ggggttcggc	ttctggcgctg	3480
tgaccggcg	ctctagagcc	tctgtctaacc	atgtttcatgc	cttcttcttt	ttcctacagc	3540
tcctggggcaa	cgtgctgggt	gttgtgctgt	ctcatattt	tggcaaagaa	ttcgccacca	3600
tggtgagcaa	gggcgaggag	ctgttcaccg	gggtggtgcc	catcctggtc	gagctggacg	3660
gcgacgtaaa	cggccacaag	ttcagcgtgt	ccggcgaggg	cgaggcgcat	gccacctacg	3720
gcaagctgac	cctgaagttc	atctgcacca	ccggcaagct	gcccgtgcc	tgcccccacc	3780
tcgtgaccac	cctgacctac	ggcgtgcagt	gcttcagccg	ctaccccagc	cacatgaagc	3840
agcacgactt	cttcaagttc	gccatgcccc	aaggctacgt	ccaggagcgc	accatcttct	3900
tcaaggacga	cggcaactac	aagaccgcgc	ccgaggtgaa	gttcgagggc	gacacccctg	3960
tgaaccgcat	cgagctgaag	ggcatcgact	tcaaggagga	cggcaacatc	ctggggcgaca	4020
agctggagta	caactacaac	agccacaacg	tctatatcat	ggccgacaag	cagaagaacg	4080
gcacaaaggt	gaacttcaag	atccgccaca	acatcgagga	cggcagcgtg	cagctcgccg	4140
accactacca	gcagaacacc	cccatcgggc	acggccccgt	gctgctgccc	gacaaccact	4200
acctgagcac	ccagtccgcc	ctgagcaaa	accccaacga	gaagcgcgat	cacatggtcc	4260
tgctggagtt	cgtgaccgcc	gccgggatca	ctctcggcac	ggacgagctg	tacaagtaag	4320
aattcactcc	tcagggtgcag	gctgcctatc	agaagggtgt	ggctgggtgtg	gccaatggcc	4380
tggctcacia	ataccactga	gatctttttc	cctctgccaa	aaattatggg	gacatcatga	4440
agccccttga	gcacttgact	tctggctaat	aaaggaaatt	tattttcatt	gcaatagtgt	4500
gttggaattt	tttgtgtctc	tcactcgga	ggacatatgg	gagggcaaat	cattttaaacc	4560
atcagaatga	gtatttggtt	tagagtttgg	caacatatgc	catatgctgg	ctgccatgaa	4620
caaagggtgg	tataaagagg	tcacagttat	atgaaacagc	cccctgctgt	ccatttctta	4680
ttccatagaa	aagccttgac	ttgaggttag	atttttttta	tattttgttt	tgtgttattt	4740
ttttctttaa	catccctaaa	attttcctta	catgttttac	tagccagatt	tttctctctc	4800
tcctgactac	tcccagtcac	agctgtccct	cttctcttat	gaagatccct	cgacctgacg	4860
cccaagcttg	catgcctgca	ggtcgactcc	cggtgagggc	cgggtaccga	gctcgaattc	4920
gtaatcatgg	tcatagctgt	ttcctgtgtg	aaattgttat	ccgctcacia	ttccacacaa	4980
catacgagcc	ggaagcataa	agtgtaaagc	ctgggggtgc	taatgagtga	gctaactcac	5040
attaatttgc	ttgcgctcac	tgcccgtttt	ccagtcggga	aacctgtcgt	gccagctgca	5100
ttaatgaatc	ggccaacgcg	cggggagagg	cggtttgcgt	attgggcgct	cttccgcttc	5160
ctcgctcact	gactcgctgc	gctcggtcgt	tcgggtgcgg	cgagcggtat	cagctcactc	5220
aaagcggtga	atacgggtat	ccacagaatc	aggggataac	gcaggaaaga	acatgtgagc	5280
aaaaggccag	caaaaggcca	ggaaaccgtaa	aaaggccgcg	ttgctggcgt	ttttccatag	5340
gctccgcccc	cctgacgagc	atcacaaaaa	tcgacgctca	agtcaagggt	ggcgaaaccc	5400
gacaggacta	taaagatacc	aggcgttttc	ccctggaagc	tccctcgtgc	gctcaccgtg	5460

-82-

```

tccgaccctg ccgcttaccg gataacctgtc cgccctttctc ccttcgggaa gcggtggcgct 5520
ttctcatagc tcacgctgta ggtatctcag ttccggtgtag gtcggttcgct ccaagctggg 5580
ctgtgtgcac gaaccccccg ttcagcccgga ccgctgcgccc ttatccggta actatcgtct 5640
tgagtccaac ccggttaagac acgacttatc gccactgggca gcagccactg gtaacaggat 5700
tagcagagcg aggtatgtag gcggtgctac agagttcttg aagtgggtggc ctaactacgg 5760
ctacactaga aggacagtat ttggtatctg cgctctgctg aagccagtta ccttcggaaa 5820
aagagttggt agctcttgat ccggcaaaaca aaccaccgct ggtagcgggtg gtttttttgt 5880
ttgcaagcag cagattacgc gcagaaaaaaa aggatctcaa gaagatcctt tgatcttttc 5940
tacgggggtct gacgctcagt ggaacgaaaaa ctacagttaa gggatttttg tcatgagatt 6000
atcaaaaagg atcttcacct agatcctttt aaattaaaaa tgaagtttta aatcaatcta 6060
aagtataatat gagtaaacct ggtctgacag ttaccaatgc ttaatcagtg aggcacctat 6120
ctcagcgatc tgtctatttc gttcatccat agttgcctga ctccccgtcg tgtagataac 6180
tacgatacgg gagggcttac catctggccc cagtgtgca atgataccgc gagacccacg 6240
ctcaccggct ccagatttat cagcaataaaa ccagccagcc ggaagggccg agcgcagaag 6300
tggtcctgca cctttatccg cctccatcca gtctattaat tgttgccggg aagctagagt 6360
aagtagttcg ccagttaata gtttgcgcaa cgttgttgcc attgtctacag gcactcgtgg 6420
gtcagcgtcg tcggttggtg tggcttcatt cagctccggt tcccaacgat caagggcagt 6480
tacatgattc cccatggtgt gcaaaaaaagc ggttagctcc ttcggtcctc cgatcgttgt 6540
cagaagtaag ttggccgcag tgttatcact catggttatg gcagcactgc ataattctct 6600
tactgtcatg ccatccgtaa gatgcttttc tgtgactggg gagtactcaa ccaagtcatt 6660
ctgagaatag tgtatgcggc gaccgagttg ctcttgcccg ggtcaatac gggataatac 6720
cgcgccacat agcagaactt taaaagtgtc catcattgga aaacgttctt cggggcgaaa 6780
actctcaagg atcttacccg tgttgagatc cagttcgatg taacccactc gtgcacccaa 6840
ctgatcttca gcatctttta ctttcaccag cgtttctggg tgagcaaaaa caggaaggca 6900
aaatgccgca aaaaagggaa taagggcgac acggaaatgt tgaatactca tactcttcct 6960
ttttcaatat tattgaagca tttatcaggg ttatttgtctc atgagcggat acatatttga 7020
atgtatttag aaaaaataac aaataggggt tccgcgcaca tttccccgaa aagtcgccac 7080
tgacgtagtt aacaaaaaaa agcccgcgca agcgggcttt attaccaagc gaagcgccat 7140
tcgccattca ggctgcgcaa ctgttgggaa gggcgatcgg tgccggcctc ttcgctatta 7200
cgccagctgg cgaaaggggg atgtgtgca aggcgattaa gttgggtaac gccagggttt 7260
tcccagtcac gacgttgtaa aacgacggcc agtccgtaat acgactcact taaggccttg 7320
actagagggg cgacggtata cagacatgat aagatacatt gatgagtttg gacaaaccac 7380
aactagaaat cagtgaaaaa aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt 7440
tgtaaccatt ataagctgca ataacaagt tggggtgggc agcatgagat agcaactcc 7500
ccccgcgctg gaggatcatc cagccggcgt cccggaaaac gatccgaag cccaaccttt 7560
catagaaggc ggcggtggaa tcgaaatctc gtagcacgtg tcagtcctgc tcctcggcca 7620
cgaagtgcac g 7631

```

<210> 117
 <211> 4615
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> p18attBZeo6XHS4 Plasmid

```

<400> 117
cagttgcccg gcccggaggc cagggcgaaac tcccccccc acggctgtct gccgatctcg 60
gtcatggccg gcccggaggc gtcccggaaag ttctgtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggcccgcg caccacacacc caggccaggg tggtgtccgg caccacctgg 180
tcctggaccg cgctgatgaa cagggtcacg tcgtcccggg ccacaccggc gaagtcgtcc 240
tccacgaagt cccgggagaa cccgagccgg ccggtccaga actcgaccgc tccggcgacg 300
tcgcgcgcgg tgagcaccgg aacggcactg gtcaacttgg gcttgcatgc ctgcatgctg 360
caagtttagt taaaaaagca ggcttcaatc ctgcagagaa gcttgcatgc ctgcatgctg 420
actctagtgg atcccccgcc ccgtatcccc caggtgtctg caggctcaaa gagcagcgag 480
aagcgttcag aggaaagcga tcccgtgcc aaggtgtctg gcccgggctg tccccgcacg 540
ctgcccggct cgggatgcgg ggggagcgcc ggaccggagc ggagccccgg gcggtctcgt 600
gctgcccccct agcgggggag ggacgtaat gggtctttgg ggggtctgtg ggggggctgt 660
ccccgtgagc ggatccgcgg ccccgatatc cccaggtgtc tgagggtca aagagcagcg 720
agaagcgttc agagaaaagc gatccccgtc caccttcccc gtgcccgggc tgtccccgca 780
cgctgccggc tccgggatgc ggggggagcg cggagagccc ggggggcttg ggggggggct 840
ctgctgcccc ctacgagggg ggcacgttaa ttacatccct ggggggcttg ggggggggct 900
gtccccgtga gcgatccgc ggccccgtat ccccagggtg cctgagcagg tctgagcagg 960
cgagaagcgt tcagagggaa gcatcccggt cccaccttcc cgtgccccgg gctgtccccg 1020
cacgctgccg gctcggggat gcggggggag cgcgggaccg gagcggagcc cccggcgggc 1080
cgctgctgcc ccctagcggg ggagggagct ggaggggctt tggggggggg 1140
ctgtccccgt gageggatcc gcggccccgt atccccagg tgtctgcagg ctcaaagagc 1200

```

```

agcgagaagc gttcagagga aagcgatccc gtgccacctt ccccgtagcc gggctgtccc 1260
cgcacgctgc cggctcgggg atgcgggggg agcgccggac cggagcggag ccccgggcgg 1320
ctcgctgctg cccctagcgc ggggagggac gtaattacat ccctgggggc tttggggggg 1380
ggctgtcccc gtgagcggat ccgcggcccc gtatccccca ggtgtctgca ggctcaaaga 1440
gcagcgagaa gcgttcagag gaaagcgatc ccgtgccacc ttccccgtgc ccgggctgtc 1500
cccgacgct gccggctcgg ggatgcgggg ggagcgccgg accggagcgg agccccgggc 1560
ggctcgctgc tgccccctag cgggggaggg acgtaattac atccctgggg gctttggggg 1620
ggggctgtcc ccgtgagcgg atccgcggcc ccgtatcccc cagggtgtctg caggctcaaa 1680
gagcagcgag aagcgttcag aggaaagcga tcccgtagcc ccttccccgt gccgggctg 1740
tccccgcacg ctgcccggctc ggggatgcgg ggggagcggc ggaccggagc ggagccccgg 1800
gcccgtcgct gctgccccct agcggggggag ggaagtaatt acatccctgg gggctttggg 1860
ggggggctgt ccccgtgagc ggatccgcgg ggtgcagga attcgtaatc atggtcatag 1920
ctgtttcctg tgtgaaattg ttatccgctc acaattccac acaacatacg agccggaagc 1980
ataaagtgt aagcctgggg tgcctaata gtagctaac tcacattaat tgcgttgccg 2040
tcactgcccg ctttccagtc gggaaacctg tctgtcccagc tgcattaatg aatcgcccaa 2100
cgcgcgggga gaggcgggtt gcgtattggg gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg 2220
ttatccacag aatcagggga taacgcagga gcggttgctg aggtggcgaa acccgacagg ctagtccgac 2280
gccaggaacc gtaaaaaggc cgcgttgctg ggtttttcc ataggctccg cccccctgac 2340
gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa ccccgacagg actataaaga 2400
taccagcgct tccccctgg aagctccctc gtgcgtctc ctgttccgac cctgcgctt 2460
accggatacc tgtccgcctt tctccctcgc ggaagcgtgg cgctttctca tagctcacgc 2520
tgtaggtatc tcagttcggg gtaggtcggt cgctccaagc tgggctgtgt gcacgaaccc 2580
cccgttcagc ccgaccgctg cgcttatcc ggtactatc gtcttgagtc caaccggta 2640
agacacgact tatcgccact ggcagcagcc actggttaac ggattagcag agcgaggtat 2700
gtaggcgggt ctacagagtt ctggaagtgg tggcctaact acggctacac tagaaggaca 2760
gtatttggtg tctgcgtctt gctgaagcca gttactctg gaaaaagagt tggtagctct 2820
tgatccggca aacaaaccac cgctggtagc ggtggttttt ttgtttgcaa gcagcagatt 2880
acgcgcagaa aaaaaggatc tcaagaagat cctttgatct tttctacggg gtctgacgt 2940
cagtggaacg aaaaactcac ttaagggatt ttggtcatga gattatcaaa aaggatcttc 3000
acctagatcc ttttaaatta aaaaatgaag tttaaatcaa tctaaagtat atatgagtaa 3060
acttggtctg acagtatacc atgtttaate agtgaggcac ctatctcagc gatctgtcta 3120
tttcgttcac ctgactcccc tgcaatgata gtcgttaga taactacgat acgggagggc 3180
ttaccatctg gccccagtgc cgcgtagacc caccgtcacc ggctccgat 3240
ttatcagcaa taaccagcc agccggaagg gccgagcgca gaagtgggtc tgcaacttta 3300
tccgcctcca tcaattgtgc cgggaagcta gagtaagtag ttcgccagtt 3360
aatagtttgc gcaacgttgt tgccattgct acaggcatcg tgggtgtcacg ctctgctgtt 3420
ggtagtgctt cattcagctc cggttcccaa cgatcaaggc gagttacatg atcccccatg 3480
ttgtgcaaaa aagcggttag ctccttcggt ttgtcagaag taagttggcc 3540
gcagtgttat cactcatggt ctccgatcgc ctcttactgt catgccatcc 3600
gtaagatgct tttctgtgac ctgcataatt ctcttactgt atagtgtatg 3660
cggcgaccga gttgctcttg tcaaccaagt cattctgaga ataccgccc 3720
actttaaaag tgcctcatcat ataccgggata ataccgccc acatagcaga 3780
ccgctgttga gatccagttc tcttcggggc gaaaactctc aaggatctta 3840
ttacttttca ccagcgtttc actcgtgcac ccaactgatc ttcagcatct 3900
ggaataaggg cgacacggaa tgggtgagca aaaaacaggaa ggcaaaatgc cgcaaaaaag 3960
agcatttatc atgttgaata tctcatgagc tcttttttca atattattga 3960
aaacaaatag ggggtccgcg ggatacatat ttgaatgtat ttagaaaaat 4020
aaaaagcccg ccgaagcggg cacatttccc cacttgatgac agttaacaaa 4080
gcaactgttg ggaagggcga ctttattacc cctcttcgct ccattcgcca ttcaggctgc 4140
ggggatgtgc tgcaaggcga ttaagtggg taacgccagg gttttcccag tcacgacgtt 4260
gtaaaacgac ggccagtccg cattgatgag gttgactaga gggctgacgg 4320
tatacagaca tgataagata tttggacaaa ccacaactag aatgcagtga 4380
aaaaaatgct ttatttgtga aatttgtgat gctattgctt tatttgtaac cattataagc 4440
tgcaataaac agttggggg gggcgaagaa ctccagcatg agatccccgc gctggaggat 4500
catccagccg cgttcccgga aaacgattcc gaagcccaac ctttcataga agggcgcggt 4560
ggaatcga aa tctcgtagca cgtgtcagtc ctgctcctcg gccacgaagt gcacg 4615

```

<210> 118

<211> 17384

<212> DNA

<213> Artificial Sequence

<220>

<223> pFK161 Plasmid

<400> 118

```

gcgacagagg gagcttccag ggggaaacgc ctggtatctt tatagtcctg tcgggggtttc 60
gccacctctg acctgagcgt cgatttttgt gatgctcgtc agggggggcgg agcctatgga 120
aaaacgccag caacgcggcc tttttacggg tcttgccctt ttgctggcct tttgctcaca 180
tggtctttcc tgcgttatcc cctgattctg tggataaccg tattaccgcc tttgagtgag 240
ctgataccgc tcgccgcagc cgaacgaccg agcgacgcga gtcagtgcgc gaggaagcgg 300
aagagcgcgt acctccgcgt ttccagactt tacgaaacac ggaaaccgaa gaccattcat 360
gttggttgctc aggtgcgcga cgttttgcag cagcagtcgc ttcacgttcg ctccgcgtatc 420
gggtgattcat tctgctaacc agtaaggcaa ccccgccagc ctagccgggt cctcaacgac 480
aggagcacga tcatgcgcac ccgtcagatc cagacatgat aagatacatt gatgagtttg 540
gacaaaccac aactagaatg cagtgaaaaa aatgctttat ttgtgaaatt tgtgatgcta 600
ttgctttatt tghtaacatt ataagctgca ataaacaagt taacaacaac aattgcattc 660
attttatggt tcaggttcag ggggaggtgt gggaggtttt ttaaagcaag taaaacctct 720
acaaatgtgg tatggctgat tatgatctct agtcaaggca ctatacatca aatattcctt 780
attaacccct ttacaaatta aaaagctaaa ggtacacaat ttttgagcat agttattaat 840
agcagacact ctatgcctgt gtggagtaag gtctttgggg atgttatgat tataactggt 900
atgcctactt ataaagggtt cagaatattt ttccataatt ttcttgata gcagtgacgc 960
tttttctttt gtggtgtaaa tagcaaagca agcaagagtt ctattactaa acacagcatg 1020
actcaaaaaa cttagcaatt ctgaaggaaa gtctttgggg tcttctacct ttctcttctt 1080
ttttggagga gtagaatggt gagagtcagc agtagcctca tcatcactag atggcatttc 1140
ttctgagcaa aacagggttt cctcattaaa ggcattccac cactgctccc attcatcagt 1200
tccataggtt ggaatctaaa atacacaaac aattagaatc agtagtttaa cacattatac 1260
acttaaaaaa tttatattta ccttagagct ttaaactctt gtaggtagtt tgtccaatta 1320
tgtcacacca cagaagtaag gttccttcac aaagatccgg accaaagcgg ccatcgtgcc 1380
tccccactcc tcaggttcgg gggcatggat cgcgggatag ccgtgctggg tttcctggat 1440
gccgacggat ttgcaactgc ggtagaactc gcgaggtcgt ccagcctcag gcagcagctg 1500
aaccaactcg cgagggggatc gagccccggg tgggccaaga actccagcat gagatccccg 1560
cgctggagga tcatccagcc ggctcccggg aaaacgattc cgaagcccaa ccttccatag 1620
aaggcggcgg tggaatcgaa atctcgtgat ggcaggttgg gcgtcgcttg gtcggtcatt 1680
tcgaacccca gagtcccgt cagaagaact cgtcaagaag gcgatagaag gcgatgcgt 1740
gcgaatcggg agcggcgata ccgtaaagca cgaggaagcg gtcagcccat tcgccgcaa 1800
gctcttcagc aatatcacgg gtacccaacg ctatgtcctg atagcgggtc gccacacca 1860
gccggccaca gtcgatgaat ccagaaaagc ggccattttc caccatgata tcgggcaagc 1920
aggcatcgcc atgggtccag acgagatcct agccgctggg atgcgcgctt tgagctggc 1980
gaacagttcg gctggcgaga gcccctgatg ctcttcgtcc agatcatcct gatcgacaag 2040
accggcttcc atccgagtac gtgctcgctc gatgcgatgt ttcgcttggt ggtcgaatgg 2100
gcaggtagcc tcatccagcg tatgcagcg atgcagggat atcctgcccc tggcaacttt 2160
ctcggcagga gcaagggtgag atgacaggag atcctgcccc ggcacttcgc ccaatagcag 2220
ccagtccttt cccgcttcag tgacaacgtc gagcacagct ggcgaaggaa cgcccgtcgt 2280
ggccagccac gatagccgag ctgctcgctc cgcaggttca ttcagggcac cggacaggct 2340
ggctctgaca aaaagaaccc ggccgccccg cgctgacagc cggaaacacg cggcatcaga 2400
gcagccgatt gtctgtttgt cccagtcata gccgaatagc ctctccaccc aagcggccgg 2460
agaactcggt tgcaatccat ctgtttcaat cgcgcgaaac gatcctcatc ctgctcttgg 2520
atcagatctt gatccccctg gccatcagat ccttgggcgg ccttgccagct aagaaagcca tccagtttac 2580
tttgaggggg tttcccaact taccagaggg ccgcccagct ggcaattccg gttcgcttgc 2640
tgtccataaa accgcccagt ctagetatcg ccatgtaagc cactgcaag ctactgctt 2700
tctcttttgg cttgctgttt ccttgttcca gatagcccag tagctgacat tcatccgggg 2760
tcagcaccgt ttctgcggac tggttttcta ttccttttagc agcccttggc 2820
cctgagtgcc ttgcggcagc gtgaaaagct tttgcaaaag cctaggcctc caaaaaagc 2880
tccctactac ttctggaata gctcagaggc cgaggcggcc taaataaaaa aaattagtc 2940
gccatggggg ggagaattgg cggaactggg cggagttagg ggcgggatgg gcggagttag 3000
gggcccgggt atggttgctg acctggttgc tgactaattg agatgcagtc tttgcatact 3060
ggagcctggg gactttccac ggggagcctg gggactttcc agatgcagtc tttgcatact 3120
tctgctgctt ggggagcctg gggactttcc acaccctaac tgacacacat tccacagccg 3180
gatctgcagg acccaacgct gcccagatg gcgcgcgtg ggctgctgga gatggcggag 3240
gcatgggata tgttctgcca agggttggtt tgccgattca cagttctccg caagaattga 3300
ttggctccaa ttcttgaggt ggtgaatccg ttagcgaggt gccgcgggt tccattcagg 3360
tcgaggttgc ccggctccat gcaccgcgag gcaacgcggg gaggcagaca aggtataggg 3420
cggcgctacc aatccatgcc aaccgcttcc atgtgctcgc cgaggcgcat aaatcgccgt 3480
gacgatcagc ggtccaatga tcgaagttag gctggttaaga gccgcgagcg atccttgaag 3540
ctgtccctga tggctgctat ctacctgctt ggaagcatg gcttgcaacg cggcatcccg 3600
atgccgcggg aagcgagaag aatcataatg ggggaaggca tccagcctcg cgtcgccgac 3660
gccagcaaga cgtagcccag cgcgtcgggc cgccatgccg gcgataatgg cctgcttctc 3720
gccgaaacgt ttggtggcgg gaccagtgac gaaggcttga gcgctccag gcgagggcgt gcaagattcc 3780
gaataccgca agcgacaggc cgatcatcgt caggtccag cgaaaagcgg cctcgccgaa 3840
aatgacccag agcgtgccc gcacctgtcc tacgagttgc atgataaaga agacagtcac 3900
aagtgcggcg acgatagtca tgccccgcgc ccaccggaag gagctgactg ggttgaaggc 3960
tctcaagggc atcggtcgac gctctccctt atgcgactcc tgcattagga agcagccag 4020

```


tagtaggttg	aggccgttga	gcaccgcgcg	cgcaaggaat	ggtgcatgca	aggagatggc	4080
gccccacagt	cccccgccca	cgggcctgcc	accataccca	cgccgaaaca	agcgctcatg	4140
agcccgaaagt	ggcgagcccg	atcttcccca	tcggtgatgt	cgccgatata	ggcgccagca	4200
accgcacctg	tggcgccggt	gatgcgggcc	acgatgcgtc	cgccgtagag	gatcttggca	4260
gtcacagcat	gcgcataatcc	atgcttcgac	catgcgtca	caaagtaggt	gaatgcgcaa	4320
tgtagtaccc	acatcgatcat	cgctttccac	tgctctcgcg	aataaagatg	gaaaaatcaat	4380
ctcatggtaa	tagtccatga	aaatccttgt	attcataaat	cctccaggta	gctatatgca	4440
aattgaaaca	aaagagatgg	tgatctttct	aagagatgat	ggaatctccc	ttcagtatcc	4500
cgatgggtcaa	tgcgctggat	atgggataga	tgggaatatg	ctgattttta	tgggacagag	4560
ttgcgaactg	ttcccaacta	aaatcatttt	gcacgatcag	cgcactacga	actttaccca	4620
caaatagtca	ggtaatgaat	cctgatataa	agacaggttg	ataaatcagt	cttctacgcg	4680
catcgcacgc	gcacaccgta	gaaagtcttt	cagttgtgag	cctgggcaaa	cogttaactt	4740
tcggcggtct	tgctgtgcga	caggctcacg	tctaaaagga	aataaatcat	gggtcataaa	4800
attatcacgt	tgctcggcgc	ggcgacggat	gttctgtatg	cgctgttttt	ccgtggcgcg	4860
ttgctgtctg	gtgatctgcc	ttctaaatct	ggcacagccg	aattgcgcga	gcttggtttt	4920
gctgaaacca	gacacacagc	aactgaatac	cagaaagaaa	atcactttac	ctttctgaca	4980
tcagaagggc	agaaatttgc	cgttgaacac	ctgggtcaata	cgctgttttg	tgagcagcaa	5040
tattgcgctt	cgatgcgctt	tggcggtgag	attgatccct	ctgctgcaca	aaaggcaatc	5100
gacgagctgg	accagcgcat	togtgacacc	gtctccttcg	aacttatctg	caatggagtg	5160
tcattcatca	aggacgcgcg	tatcgcaaat	gggtgctatcc	acgcagcggc	aatcgaaaca	5220
cctcgcgcgg	tgaccaatat	ctacaacatc	agccttggta	tccagcgtga	tgagccagcg	5280
cagaacaagg	taaccgtcag	tgccgataag	ttcaaagtta	aacctgggtg	tgataccaac	5340
attgaaacgt	tgatcgaaaa	cgcgctgaaa	aacgctgctg	aatgtgcggc	gctggatgtc	5400
acaaagcaaa	tggcagcaga	caagaaagcg	atggatgaac	tggcttccta	tgcccgacag	5460
gccatcatga	tggaaatgttt	ccccgggtgg	gttatctggc	agcagtgccg	tcgatagtat	5520
gcaattgata	attattatca	tttgcggtgc	ctttccggcg	atccgccttg	ttacggggcg	5580
gcgacctcgc	gggttttcgc	tatttatgaa	aattttccgg	tttaaggcgt	ttccgttctt	5640
cttcgctcata	acttaatgtt	tttattttaa	ataccctctg	aaaagaaagg	aaacgacagg	5700
tgtctgaaagc	gagctttttg	gcctctgtgc	tttcccttct	ctgtttttgt	ccgtggaaatg	5760
aacaatggaa	gtcaacaaaa	agcagctggc	tgacattttc	ggtgcgagta	tcogtaccat	5820
tcagaactgg	caggaacagg	gaatgcccgt	tctgcgaggg	ggtggcaagg	gtaatgaggt	5880
gctttatgac	tctgcgcgcg	tcataaaaatg	gtatgccgaa	agggatgctg	aaattgagaa	5940
cgaaagcaaa	cgcgggagg	gcggcagccc	gcggcagccc	agcgaggcag	atccacagga	6000
cggtgtgggt	cgccatgatc	gcgtagtcga	tagtggctcc	aagtagcgaa	gcgagcagga	6060
ctgggcggcg	gcgaagcggg	cggacagtcg	tccgagaacg	ggtgcgcata	gaaattgcac	6120
caacgcatac	agcgtagca	gcacgccata	gtgctggcgg	atgctgtcgg	aatggacgat	6180
atcccgcaag	aggcccgcca	gtaccggcat	aaccaagcct	atgcctacag	catccagggt	6240
gacggtgcgc	aggatgacga	tgagcgcat	gttagatttc	atacacgggt	cctgactcgc	6300
ttagcattga	aactgtgata	aactaccgca	ttaaagctta	tcgatgataa	gcgggtcaaa	6360
atgagaatcc	gcggcgctcc	ttctcgttct	gccagcgggc	cctcgtctct	ccaccccatc	6420
cgtctgcccg	tggtgtgtgg	aaggcagggg	tgcggtctct	cggcccgcag	ctgccccgcg	6480
cgcacttttc	tcagtggttc	gcgtggctcc	tgtggatgtg	tgaggcgccc	ggttgtgccc	6540
tcacgtgttt	cactttgggtc	gtgtctcgct	tgaccattgt	cccagagtcg	gtggatgtgg	6600
ccggtggcgt	tgcataccct	tcccgtctgg	tgtgtgcacg	cgctgtttct	tgtaagcgtc	6660
gaggtgctcc	tggagcgttc	cagggtttgtc	tcttaggtgc	ctgcttctga	gctgggtggg	6720
gcgctcccca	ttccctgggtg	tgccctcggg	gctccgtctg	gctgtgtgcc	ttcccggttg	6780
tgtctgagaa	gcccgtgaga	gggggggtcg	ggagagaagg	aggggcaaga	cccccttctc	6840
tcgtcgggtg	aggcgcccac	cccgcgacta	gtacgcctgt	gcgtagggct	ggtgctgagc	6900
ggtcgcggct	gggggtggaa	agtttctcga	gagactcatt	gctttcccg	ggggagcttt	6960
gagaggcctg	gctttcgggg	gggaccgggt	gcagggtctc	ccctgtccgc	ggatgctcag	7020
aatgcccttg	gaagagaacc	ttcctgttgc	cgcagacccc	cccgcgcggg	cgcccgcgtg	7080
ttggtctctc	gggttccctg	tgtgctcgtc	gcactgcattc	tctctcggtg	gcccgggctc	7140
gtcgggggtt	tgggtccgctc	ccgcctcag	tgagaaagt	tccttctcta	gctatcttcc	7200
ggaaagggtg	cgggcttctc	acggtctcga	gggtctctc	ccgaatggtc	ccctggaggg	7260
ctcgcctctc	gacgcctcc	cgcgcgcgca	gcgtttgctc	tctcgtctac	cgcggcccg	7320
ggcctccccg	ctccgagttc	ggggagggat	cacgcggggc	agagcctgtc	tgtcgtcctg	7380
ccgttgctgc	ggagcatgtg	gctcggcttg	tgtggttggg	ggctggggag	agggtccggt	7440
gcacaccccc	gcagtcgcgt	actttccctc	cctcctgagg	gcgcgcgtgc	ggagcgggtg	7500
tgggtaggcg	acgggtgggtc	cccgggtccc	cacccgtctt	cccgtgcctc	acccgtgcct	7560
tcgctcgctc	gcgtccctct	cgtcgcgtc	cacgactttg	gccgctcccg	cgacggcgcc	7620
ctgcgcgcgc	cgttggtcgct	gctgtgtgct	tctcggtgtg	tgtggttgtg	tcgctcgcgc	7680
cccccttctc	cgcggcagcg	ttccacgggc	tggcgaaatc	gcgggagtc	tccttccctc	7740
cctcgggggtc	gagagggtcc	gtgtctggcg	ttgattgatc	tcgctctcgg	ggacggggac	7800
gttctgtggg	agaacggctg	ttggccgcgt	ccggcgcgac	gtcggacgtg	gggaaccact	7860
gccgctcggg	ggtcttcgctc	ggtaggcatc	ggtgtgtcgg	catcggtctc	tctctcgtgt	7920
cgggtgcgcg	tcctcgggct	cccggggggc	cgtcgtgttt	cgggtcgggt	cgggcgtgca	7980
ggtgtggtgg	gactgctcag	gggagtgggtg	cagtgatgat	ccgcgcgggt	ttgcctcgcg	8040

tgccctgacc	gggtccgacgc	ccgagcggtc	tctcggtccc	ttgtgaggac	ccccctccgg	8100
gaggggccc	tttcggccgc	ccttgccgtc	gtcgccggcc	ctcgttctgc	tgtgtcgttc	8160
ccccctccc	gctcgccgca	gccggtcttt	tttctctct	ccccctct	cctctgactg	8220
accggtggcc	gtgtgttcgg	accccccgca	tgggggcggc	cgggcacgta	cgcgccggg	8280
cggtcacccg	gggtcttggg	gggggcccag	gggtaagaaa	gtcggctcgg	cgggcgagg	8340
gagctgtggt	ttggagggcg	tcccggcccc	gcggccgtgg	cggtgtcttg	cgcggtcttg	8400
gagagggctg	cgtgcgaggg	gaaaaggttg	ccccgcgagg	gcaaagggaa	agaggctagc	8460
agtggctcatt	gtcccagcgg	tgtgggtggt	tgttggccga	gggtgcgtctg	gggggctcgt	8520
ccggccctgt	cgtccgtcgg	gaaggcgctg	gttggggcct	gccggagtgc	cgagggtggg	8580
accctggcgg	tgggattaac	cccgccgcgg	tgtcccgggtg	tggcggtggg	ggctccggtc	8640
gatgtctacc	tccctctccc	cgaggtctca	ggccttctcc	gcgcgggctc	tcggccctcc	8700
cctcgttcct	ccctctcgcg	gggttcaagt	cgtcgctcga	cctcccctcc	tcctctcttc	8760
catctctcgc	gcaatggcgc	cgccccaggt	cacgggtggg	tcgtcctccg	cctccgcttc	8820
tcgcgggggg	ctggccgctg	tcgggtctct	cctgcccagc	ccccgttggc	gtgggtctct	8880
ctcgccggct	ctcgccgactc	ctgggtctcg	ccggagggtc	agggggcttc	ccgggtcccc	8940
gacgttgccg	ctcgctgctg	tgtgtctggg	gggggcccgc	tgccggcctc	gcccgcctcg	9000
gagccctcgc	cgcacccgcg	gggtgtcggt	ttcgccgcgc	ggtcagttgg	gcccgtggcg	9060
tgtgtcgcgt	cgggagcgtg	tcgcgctcgc	ggcggttaga	cgcggtgtgc	gccgggctcc	9120
gacgggtggc	ctatccaggg	ctcgcccccg	ccgacccccg	cctgcccgtc	ccggtgggtg	9180
tcgttgggtg	ggggagtgaa	tggtgtctacc	gggtcattccc	tcccgcgtgg	tttgactgtc	9240
tcgcccgtgt	cgcgcttctc	tttcgcgcaa	ccccacgccc	aaccaccac	cctgtctctc	9300
cgcccccggg	cggtcgacgt	tcgggtctct	ccgatgccga	ggggttcggg	atttgtgccg	9360
gggacggagg	ggagagcggg	taagagaggt	gtcggagagc	tgtcccgggg	cgacgctcgg	9420
gttggctttg	ccgcgtgcgt	gtgctcgcg	acgggttttg	tcggaccccc	acggggtcgg	9480
tcggcccgca	tgcactctcc	cggtccgcgc	gagcgcgcgc	ccggctcacc	cccggtttgt	9540
cctcccgcca	ggctctccgc	cgccgcgcgc	tcctctctct	ctctcgcgct	ctctgtcccc	9600
cctggtcctg	ccccacccc	gacgtccgc	tcgcgcttcc	ttacctggtt	gatcctgcc	9660
ggtagcatat	gcttgtctca	aagattaagc	catgcatgtc	taagtacgca	cggccgggtac	9720
agtgaactg	cgaatggctc	attaaatcag	ttatgggtcc	tttgggtcgt	cgctcctctc	9780
ctacttggat	aactgttgta	attctagagc	taatacatgc	cgacggcgcc	tgaccctctc	9840
tcgggggggg	ggatgctgct	atttatcaga	tcaaaaccaa	cccggtgagc	tcctcccg	9900
ctccggccgg	gggtcgggcg	ccggcggtt	gggtactcta	gataacctcg	ggccgatcgc	9960
acgccccccg	tgccggcgac	gacctattcg	aacgtctgcc	ctatcaactt	tcgatggtag	10020
tcgcccgtgc	taccatgggtg	accacgggtg	acggggaatc	aggggttcgat	tcgggagagg	10080
gagcctgaga	aacggctacc	acatccaagg	aaggcagcag	gcgcgcaaat	taccactcc	10140
cgacccgggg	aggtagtac	gaaaaataac	aatacaggac	tctttcgagg	ccctctgaatt	10200
ggaatgagtc	cactttaaat	cctttaacga	ggatccattg	gagggcaagt	ctggtgccag	10260
cagccgcggg	aattccagct	ccaatagcgt	atattaaagt	tgctgcagtt	aaaaagctcg	10320
tagttggatc	ttgggagcgg	gcggcggtc	cgccgcgagg	cgagtaccg	ccgctcccc	10380
ccccttgcc	ctcgccgcgc	cctcgatgct	cttagctgag	tgtcccgcgg	ggcccgagc	10440
gtttactttg	aaaaaattag	agtgttcaaa	cgaggccgga	gccgcctgga	taccgagct	10500
aggaataatg	taataaggacc	gcggttctat	ttgtttggtt	ttcggaactg	agggcatgat	10560
taagagggac	ggccgggggg	attcgtattg	cgccgctaga	gggtgaaattc	ttggaccggc	10620
gcaagacgga	ccagagcgaa	agcatttgcc	aagaatgttt	tcattaatca	agaacgaaag	10680
tcggaggttc	gaagacgatc	agataccgtc	gtagttccga	ccataaacga	tgccgactgg	10740
cgatgcggcg	gcgttattcc	catgaccgcg	cgggcagctt	ccgggaaacc	aaagtctttg	10800
ggttccgggg	ggagtatggg	tgcaaagctg	aaacttaaa	gaattgacgg	aagggcacca	10860
ccaggagtgg	gcctgcggct	taatttgact	caacacggga	aacctcacc	ggcccgaca	10920
cgacagggat	tgacagattg	atagctcttt	ctcgattccg	tgggtgggtg	tgcatggccg	10980
ttcttagttg	gtggagcgat	ttgtctgggt	aattccgata	acgaacgaga	ctctggcatg	11040
ctaactagtt	acgcgacccc	cgagcggtcg	gcgtccccc	acttcttaga	gggacaagtg	11100
gcgttcagcc	acccgagatt	gagcaataac	aggtctgtga	tgcccttaga	tgtccggggc	11160
tgcacgcgcg	ctacactgac	tggtcagcgg	tgtgcctacc	ctgcgcgggc	aggcgcggtg	11220
aaaccgttga	acccattcgg	tgatggggat	cggggattgc	aattattccc	catgaacgag	11280
gaattccccag	taagtgcggg	tcataagctt	gcgttgatta	agtccttgcc	ctttgtacac	11340
acgcgccgct	gctactaccg	attggatggg	ttagtgggg	cctcggtatcg	gccccgcggg	11400
ggtcggccca	cgcccttggc	ggagcgctga	gaagacggtc	gaacttgact	atctagagga	11460
agtaaaagtc	ttccgtaggt	ttccgtaggt	aagacctgcg	aaagatcatt	aaacgggaga	11520
ctgtggaggga	gcggcggtcg	ggcccgctct	ccccgctctg	tgtgtgtcct	cgccgggagg	11580
cgcgctgcgc	ccgggtcccc	tcgcccgcgt	gtggagcgag	gtgtctggag	tgaggtgaga	11640
gaaggggtgg	gtgggtcggg	tctgggtccg	tctgggacgg	cctccgattt	cccccccc	11700
tccctctctc	ctcgtegggc	tctgacctcg	ccacctacc	gcggcgccgg	ctgctcgccg	11760
gcgtcttgcc	tcttcccgct	ccggtctctc	cgtgtctacg	aggggctgta	cgtcgttacg	11820
gggtttttgac	ccgtccgggg	ggcggtcggg	cgcgctttgc	cgcgctttgc	tctccgggca	11880
cccccccccg	ccgcggctct	ggcttttctc	cgttggctgg	ggcggtttgc	gcgtgtgggg	11940
ggatgtgag	gtcgcgtgtg	ggctcgcccc	tcccgatgcc	acgctttctc	ggcctcgctg	12000
gtcctccccg	ctcctgtccc	gggtacctag	ctgtcgctgt	ccggcgcgga	ggtttaagga	12060

ccccgggggg	gtcgccctgc	cgccccccagg	gtcgggggggc	gggtggggccc	gtaggggaagt	12120
cgggtcggtcg	ggcggtctctc	cctcagactc	catgacctctc	ctccccccgc	tgccgcggtt	12180
cccagggcgg	cgggtcggtg	gggggggtgga	tgtctggagc	cccctcgggc	gccgtggggg	12240
cccagaccgc	ggcgccggct	tgcccgattt	ccgcgggtcg	gtcctgtcgg	tgccggtcgt	12300
gggttcccg	gtcgttcccg	tgtttttccg	ctcccgacc	tttttttttc	ctcccccca	12360
cacgtgtctc	gtttcgttcc	tgctggccgg	cctgaggcta	cccctcggtc	catctgttct	12420
cctctctctc	cggggagagg	agggcggttg	tcggtggggg	actgtgccgt	cgtcagcacc	12480
cgtgagttcg	ctcacaccgc	aaataccgat	acgactetta	gcgggtggatc	actcggctcg	12540
tgctgcgatg	aagaacgcag	ctagctgcga	gaattaatgt	gaattgcagg	acacattgat	12600
catcgacact	tcgaacgcac	ttgcggcccc	gggttcctcc	cggggctacg	cctgtctgag	12660
cgtcgggtga	cgatcaatcg	cgtcacccgc	tgccggtgggt	gctgcgcggc	tgggagtttg	12720
ctcgcagggc	caacccccca	accgggtcg	ggccctccgt	ctcccgaaat	tcagacgtgt	12780
gggcgggtgt	cgggtgtggcg	cgcgcgcccg	cgtcgcggag	cctgggtctcc	cccgcgcac	12840
gcgcgtcgcg	gcttcttccc	gctccgcgct	tcccgccctc	gcccgtgcac	cccgggtcgt	12900
gcctcgcgtc	ggcgccctcc	ggaccgctgc	ctcaccagtc	tttctcggtc	ccgtgccccg	12960
tggaacccca	ccgcgcccc	gtggcgcccg	ggggtggggc	cgtccgcac	tgctctggtc	13020
gaggttggcg	ggtgaggggtg	tgctgtgcgc	gaggtgggtg	tcgggtccct	gcgccgcggg	13080
gggtgtcggg	gtggcggtcg	acgagggcg	gtcgttcgcc	tgccggtggt	gtctgtgtgt	13140
ggttgggtct	tgcgctgggg	gagggcggggt	cgaccgctcg	cggggttggc	gcggtcgccc	13200
ggcgccgcgc	accctccggc	ttgtgtggag	ggagagcgag	ggcgagaacg	gagagaggtg	13260
gtatccccgc	tgccgttcgc	agggaggggt	tgccgtcccg	cgtccgtccg	tccctccctc	13320
cctcgggtggg	cgcccttcgcg	ccgcacgcgg	ccgctagggg	cggtcggggc	ccgtggcccc	13380
cgtggctctt	cttcgtctcc	gcttctcctt	caccggggcg	gtaccgcctc	cggcgcggcg	13440
ccgcgggacg	ccgcggcgctc	cgtgcgcgta	tgccagtcac	ccccgggtgt	tgccagttcg	13500
gggagggaga	gggcctcgct	gaccggttgc	gtcccggtt	ccctgggggg	gacccggtcg	13560
ctgttgggctg	tgccgtcccg	gggttgctgt	tgagtaagat	cctccacccc	cgcccgccctc	13620
ccctccccgc	ggcctctcgg	ggaccctctg	agacggttct	ccggctcgtc	ctcccggtcc	13680
gcccgggtgcc	gtctctttcc	cgcccgccctc	ctcgtctctc	tcttcccgcg	gctggggcgcg	13740
tgccccccct	ttctgaccgc	gacctcagat	cagacgtggc	gaccgcgtga	atttaagcat	13800
attagtcagc	ggaggaaaag	aaactaacca	ggattccctc	agtaacggcg	agtgaacag	13860
gaagagccca	gcgcggaatc	ccgcgcgcgc	gtcgcggcgt	gggaaatgtg	gcgtacggaa	13920
gaccacactcc	ccggcgccgc	tcgtgggggg	cccaagtcct	tctgatcgag	gcccagcccc	13980
tggaacgggtg	gagggccggtg	gcggcccgcg	cgccgcggcg	tcgggtcttc	ccggagtcgg	14040
gltgtctggg	aatgcagccc	aaagcggggtg	gtaaactcca	tctaaggcta	aataccggca	14100
tcagaccgat	agtcaacaag	taccgtaagg	gaaagtgtga	aagaactttg	aagagagagt	14160
cgaaaggggc	gtgaaaccgt	taagaggtaa	acgggtgggg	tccgcgcagt	ccgcccggag	14220
gattcaacccc	ggcggcgcgc	gtccggccgt	gcccgggtgt	cccggcggtat	ctttcccgct	14280
ccccgttccct	cccgacccct	ccaccgcgcg	gtcgttcccc	tcttccctcc	cgcgccggcg	14340
gctccggcg	gcgggcgcgc	gggggtgggtg	gggtgtggcg	cgccggcggg	gcccgggggtg	14400
gggtcggcgg	gggaccgccc	ccggccgcgc	accggccgcc	gcccggcgca	cttccaccgt	14460
ggcggtgcgc	cgcgaccggc	tccgggacgc	ccgggaaggc	ccggtgggga	aggtggctcg	14520
ggggggggcgg	cgcgctctcag	ggcgcgccga	accacctcac	cccagtggtt	acagccctcc	14580
ggcgccgctt	tcggcgaaatc	ccggggccga	taaccgctcg	taccgctcgc	cgcgctctcc	14640
ctctcccccc	gtccgcctcc	cgggcgggcg	tgggggtggg	ggccggggcg	cccctcccac	14700
ggcgcgaccg	ctctcccacc	cccctccgtc	gcctctctcg	gggcccgggtg	ggggggcgggg	14760
cggactgtcc	ccagtgcgcc	ccggcgctcg	tcgcgcgctc	gggtcccggg	gggaccgtcg	14820
gtcacgcgctc	tcccgcagaa	gccgagcgca	cggggtcggc	ggcgatgtcg	gctaccacc	14880
cgaccgcgtct	tgaaacacgg	accaaggagt	ctaaccgctg	cgcgagttag	gggctcgtcc	14940
gaaagccgcc	gtggcgcaat	gaaggtgaag	ggccccgccc	ggggggccga	ggtgggatcc	15000
cgaggccctct	ccagtccgcc	gagggcgcac	caccggcccc	tctcgccccg	cgcgccgggg	15060
aggtggagca	cgagcgtacg	cgttaggacc	cgaaagatgg	tgaactatgc	ttgggcaggg	15120
cgaagccaga	ggaaaactctg	gtggaggtcc	gtagcggtcc	tgacgtgcaa	atcggtcgtc	15180
cgacctgggt	ataggggcga	aagactaatc	gaaccatcta	gtagctgggt	ccctccgaag	15240
tttccctcag	gatagctggc	gctctcgctc	ccgacgtacg	cagtttttatc	cggtaaagcg	15300
aatgattaga	ggtcttgggg	ccgaaacgat	ctcaacctat	tctcaaaactt	taaatgggta	15360
agaagcccgg	ctcgtctggcg	tggagccggg	cgtggaatgc	gagtgcctag	tgggccactt	15420
ttggttaagca	gaactggcgc	tgccgggatga	accgaacgcc	gggttaaggc	gcccgatgcc	15480
gacgctcatc	agaccccaga	aaaggtgttg	gttgatatag	acagcaggac	ggtggccatg	15540
gaagtccgaa	tccgctaagg	agtgtgtaac	aactcacctg	ccgaatcaac	tagccctgaa	15600
aatggtggc	gctggagcgt	cgggcccata	cccgcccgctc	gccgcagtcg	gaaccggaacg	15660
ggacgggagc	ggcgcggaat	tcttgaagac	gaaagggcct	cgtgatacgc	ctatttttat	15720
aggttaatgt	catgataata	atggtttctt	agacgtcagg	tggcactttt	cggggaaatg	15780
tgccgggaac	ccctatttgt	ttatttttct	aaatcacctc	aaatatgtat	ccgctcatga	15840
gacaataacc	cttcaataat	cttcaataat	attgaaaaag	gaagagtatg	agtatccaac	15900
atttccgtgt	cgcccttatt	cccttttttg	cggcattttg	cttcctgttt	ttgctcacc	15960
agaaacgctg	gtgaaagtaa	aagatgctga	agatcagttg	ggtgcacgag	tggtttacat	16020
cgaactggat	ctcaacagcg	gtaagatcct	tgagagtttt	cgccccgaag	aacgttttcc	16080

-88-

```

aatgatgagc actttttaag ttctgctatg tggcgcggtg ttatcccggtg ttgacgcggg 16140
gcaagagcaa ctoggtcgcc gcatacacta ttctcagaat gacttggttg agtactcacc 16200
agtacagaa aagcatctta cggatggcat gacagtaaga gaattatgca gtgctgccat 16260
aaccatgagt gataacactg cggccaactt acttctgaca acgatcggag gaccgaagga 16320
gctaaccgct tttttgcaca acatggggga tcatgttaact cgcttgatc gttgggaacc 16380
ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg cagcaatggc 16440
aacaacgttg cgcaaacat taactggcga actacttact ctagcttccc ggcaacaatt 16500
aatagactgg atggaggcgg ataaagtgtc aggaccactt ctgcgtcgg cccttccggc 16560
tggctgggtt attgctgata aatctggagc cggtgagcgt gggctcgcg gtatcattgc 16620
agcactgggg ccagatggta agccctcccg tatcgtagtt atctacacga cggggagtc 16680
ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac tgattaagca 16740
ttggtaactg tcagaccaag ttactcata tatactttag attgatttaa aacttcattt 16800
ttaatttaaa aggatctagg tgaagatcct ttttgataat ctcatgacca aaatccctta 16860
acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag gatcttcttg 16920
agatcctttt ttctgcgctg taatctgctg cttgcaaaac aaaaaaccac cgctaccagg 16980
ggtggtttgt ttgccggatc aagagctacc aactcttttt ccgaaggtaa ctggcttcag 17040
cagagcgcag ataccaata ctgtccttct agtgtagccg tagttaggcc accacttcaa 17100
gaactctgta gcaccgccta catacctcgc tctgctaate ctgttaccag tgggtgctgc 17160
cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac cggataaggg 17220
gcagcgggtc ggctgaacgg ggggttcgtg cacacagccc agcttggagc gaacgacct 17280
caccgaactg agatacctac agcgtgagct atgagaaagc gccacgcttc cgaagggaga 17340
aaggcggaca ggtatccggt aagcggcagg gtcggaacag gaga 17384

```

<210> 119

<211> 2814

<212> DNA

<213> Artificial Sequence

<220>

<223> pLITMUS38 Plasmid

<400> 119

```

gttaactacg tcagggtggca cttttcgggg aaatgtgcgc ggaaccoccta tttgtttatt 60
tttctaataa cattcaaata tgtatccgct catgagacaa taaccctgat aaatgcttca 120
ataatattga aaaaggaaga gtatgagtat tcaacatttc cgtgtcgccc ttattccctt 180
ttttgcggca ttttgccttc ctgtttttgc tcacccagaa acgctggtga aagtaaaaga 240
tgctgaagat cagttgggtg cacgagtggtg ttacatcgaa ctggatctca acagcggtaa 300
gatccttgag agttttcgcc ccgaagaacg ttctccaatg atgagcactt taaagttct 360
gctatgtggc gcggtattat cccgtgttga cgccgggcaa gagcaactcg gtcgcgcgat 420
acactattct cagaatgact tgggtgagta ctcaccagtc acagaaaagc atcttacgga 480
tggcatgaca gtaagagaat tatgcagtgc tgccataacc atgagtgata acactgcggc 540
caacttactt ctgacaacga tcggaggacc gaaggagcta accgcttttt tgcacaacat 600
gggggcatcat gtaactcgcc ttgatcgttg ggaaccggag ctgaatgaag ccataccaaa 660
cgacgagcgt gacaccaaga tgccctgtagc aatggcaaca acgttgcgca aactattaac 720
tggcgaacta ctactctag cttcccggca acaattaata gactggatgg aggcggataa 780
agttgcagga ccacttctgc gctcgccctc tccggctggc tggtttattg ctgataaatc 840
tggagccggg gagcgtgggt ctgcgggtat cattgcagca ctggggccag atggtaagcc 900
ctcccgatc gtatgtatct acacgacggg gagtcaggca actatggatg aacgaaatag 960
acagatcgct gagatagggt cctcaactgat taagcattgg taactgtcag accaagttta 1020
ctcatatata ctttagattg atttaccocg gttgataatc agaaaagccc caaaaacagg 1080
aagattgtat aagcaaatat ttaaattgta aacgttaata ttttgtaaa attcgcgtta 1140
aatttttgtt aaatcagctc attttttaac caataggccg aaatcggcaa aatcccttat 1200
aaatcaaaag aatagcccgga gatagggttg agtggtgttc cagtttggaa caagagtcca 1260
ctattaaaga acgtggactc caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc 1320
ccactacgtg aaccatcacc caaatcaagt tttttggggg cgaggtgccg taaagcacta 1380
aatcggaacc ctaaagggag cccccgattt agagcttgac ggggaaagcg aacgtggcga 1440
gaaaggaagg gaagaaaagg aaaggagcgg gcgctagggc gctggcaagt gtacgggtca 1500
cgctgcgcgt aaccaccaca cccgcgcgcg ttaactgcgc gctacagggc gctgaaaagg 1560
atctagggtg agatcctttt tgataaatctc atgacaaaaa tcccttaacg tgagttttcg 1620
ttccactgag cgtcagaccc cgtagaaaaa atcaaaggat cttcttgaga tctttttttt 1680
ctgcgcgtaa tctgtgctt gcaaacaaaa taccagcggt taccagcggt ggtttgtttg 1740
ccggatcaag agctaccaac tctttttccg aaggtaactg gcttcagcag agcgcagata 1800
ccaaatactg tttctctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca 1860
ccgctacat acctcgctc gctaactcgt ttaccagtg ctgctgccag tgggtgataag 1920
tcgtgtctta ccgggttgga ctcaagacga tagttaccgg ataaggcgca gcggctggggc 1980
tgaacggggg gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga 2040
tacctacagc gtgagctatg agaaagcgcc acgcttcccg aaggggagaaa ggcggagacag 2100

```

-89-

tatccggtaa	gcggcagggg	cgggaacagga	gagcgcacga	gggagcttcc	aggggggaaac	2160
gcctgggtatc	tttatagttcc	tgctcgggttt	cgccacctct	gacttgagcg	tcgattttttg	2220
tgatgctcgt	cagggggggcg	gagcctatgg	aaaaacgccca	gcaacgcggc	cttttttacgg	2280
ttcctggcct	tttgctggcc	ttttgctcac	atgtaatgtg	agttagctca	ctcatttaggc	2340
accccaggct	ttacacttta	tgcttccggc	tcgtatgttg	tgtggaattg	tgagcggata	2400
acaatttcac	acaggaaaca	gctatgacca	tgattacgcc	aagctacgta	atacgactca	2460
ctagtggggc	ccgtgcaatt	gaagccgggt	ggcgccaagc	ttctctgcag	gatatactgga	2520
tccacgaatt	ggctagcttc	ggcgtgacg	cgtctccgga	tgtacaggca	tgcgctcgacc	2580
ctctagctcaa	ggccttaagt	gagtcgtatt	acggactggc	cgctcgttta	caacgtcgtg	2640
actgggaaaa	ccctggcggt	acccaactta	atcgccctgc	agcacatccc	cctttcgcca	2700
gctggcgtaa	tagcgaagag	gcccgcaccg	atcgcccttc	ccaacagttg	cgcagcctga	2760
atggcgaagt	gcgcttcgct	tggttaataaa	gcccgcttcg	gcgggctttt	tttt	2814

<210> 120

<211> 2847

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attB Plasmid

<400> 120

gttaactacg	tcaggtggca	cttttcgggg	aaatgtgcgc	ggaaccccta	tttgtttatt	60
tttctaataa	cattcaaata	tgatccgct	catgagacaa	taaccctgat	aaatgcttca	120
ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttattccctt	180
ttttgcgcca	ttttgccttc	ctgtttttgc	tcaccagaaa	acgctgggtga	aagtaaaaga	240
tgctgaagat	cagttgggtg	cacgagtggg	ttacatcgaa	ctggatctca	acagcggtaa	300
gatccttgag	agttttcgcc	ccgaagaacg	ttctccaatg	atgagcactt	ttaaagttct	360
gctatgtggc	gcggtattat	cccgtgttga	cgccggggcaa	gagcaactcg	gtcgccgcac	420
acactattct	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	480
tgccatgaca	gtaagagaat	tatgcagtgc	tgccataacc	atgagtata	acactgcggc	540
caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	tgcaacaacat	600
gggggagcat	gtaactcgcc	ttgatcgttg	ggaaccggag	ctgaatgaag	ccataccaaa	660
cgacgagcgt	gacaccacga	tgctgtagc	aatggcaaca	acgttgcgca	aactattaac	720
tgccgaacta	cttactctag	cttcccggca	acaattaata	gactggatgg	agggcgataa	780
agttgcagga	ccacttctgc	gctcggccct	tcggctggc	tggtttattg	ctgataaatc	840
tgagccgggt	gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	atggtaagcc	900
ctcccgtatc	gtagttatct	acacgcaggg	gagtcaggca	actatggatg	aacgaaatag	960
acagatcgct	gagatagggt	cctcactgat	taagcattgg	taactgtcag	accaagttta	1020
ctcatatata	cttttagattg	atttaccccg	gttgataatc	agaaaagccc	caaaaacagg	1080
aagattgtat	aagcaaatat	ttaaattgta	aacgttaata	ttttgttaaa	attcgcgtta	1140
aatttttgtt	aaatcagctc	attttttaac	caataggccg	aaatcggcaa	aatcccttat	1200
aatcaaaaag	aatagcccga	gataggggtg	agtgttgttc	cagtttgga	caagagtcca	1260
ctattaaaaga	acgtggactc	caacgtcaaa	ggcgcaaaaa	ccgtctatca	gggcgatggc	1320
ccactacgtg	aacctaccac	caaatcaagt	ttttgggggt	cgagggtgccc	taaagcacta	1380
aatcggaaacc	ctaaaggggag	cccccgattt	agagcttgac	gggggaaagcg	aacgtggcga	1440
gaaaggaagg	gaagaaaagcg	aaaggagcgg	gcgctagggc	gctggcaagt	gtagcggcca	1500
cgctgcgcgt	aaccaccaca	cccgcgcgcg	ttaatgcgcg	gctacagggc	gcgtaaaagg	1560
atctagggtga	agatccctttt	tgataatctc	atgacccaaa	tcctttaacg	tgagttttcg	1620
ttccactgag	cgtcagaccc	cgtagaaaaag	atcaaaaggat	cttcttgaga	tccttttttt	1680
ctgcgcgtaa	tctgctgctt	gcaaacaaaa	aaaccaccgc	taccagcggg	ggtttgtttg	1740
ccggatcaag	agctaccaac	tcctttttccg	aaggtaactg	gcttcagcag	agcgcagata	1800
ccaaataactg	ttcttctagt	gtagccgtag	ttaggccacc	acttcaagaa	ctctgtagca	1860
ccgcctacat	acctcgctct	gctaactctg	ttaccagtgg	ctgctgccag	tgccgataag	1920
tcgtgtctta	ccgggttgga	ctcaagacga	tagttaccgg	ataaggcgca	gcggctggggc	1980
tgaacggggg	gttcgtgcac	acagcccagc	ttggagcgaa	cgacctacac	cgaactgaga	2040
tacctacagc	gtgagctatg	agaaagcgcc	acgcttcccg	aagggagaaa	ggcggacagg	2100
tatccggtaa	gcggcagggg	cgggaacagga	gagcgcacga	gggagcttcc	aggggggaaac	2160
gcctgggtatc	tttatagttcc	tgctcgggttt	cgccacctct	gacttgagcg	tcgattttttg	2220
tgatgctcgt	cagggggggcg	gagcctatgg	aaaaacgccca	gcaacgcggc	cttttttacgg	2280
ttcctggcct	tttgctggcc	ttttgctcac	atgtaatgtg	agttagctca	ctcatttaggc	2340
accccaggct	ttacacttta	tgcttccggc	tcgtatgttg	tgtggaattg	tgagcggata	2400
acaatttcac	acaggaaaca	gctatgacca	tgattacgcc	aagctacgta	atacgactca	2460
ctagtggggc	ccgtgcaatt	gaagccgggt	ggcgccaagc	ttctctgcag	gattgaagcc	2520
tgcttttttta	tactaaacttg	agcgaatct	ggatccacga	attcgctagc	ttcgcccggt	2580
acgcgtctcc	ggatgtacag	gcctgcgtcg	accctctagt	caaggcctta	agtgagctgt	2640
attacggact	ggcgcgtcgtt	ttacaacgtc	gtgactggga	aaaccctggc	gttaccacaac	2700

-90-

ttaatcgccct	tgcagcacat	ccccctttcg	ccagctggcg	taatagcgaa	gaggcccgca	2760
ccgatcgccc	ttcccaacag	ttgcgcagcc	tgaatggcga	atggcgcttc	gcttggtaat	2820
aaagcccgct	tcggcgggct	ttttttt				2847

<210> 121

<211> 4223

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attBBSRpolyA2 Plasmid

<400> 121

accatgaaaa	cattttaacat	ttctcaacaa	gatctagaat	tagtagaagt	agcgacagag	60
aagattacaa	tgctttatga	ggataataaaa	catcatgtgg	gagcggcaat	tcgtacgaaa	120
acaggagaaa	tcatttcggc	agtacatatt	gaagcgtata	taggacgagt	aactgtttgt	180
gcagaagcca	ttgcgattgg	tagtgacgtt	tcgaatggac	aaaaggattt	tgacacgatt	240
gtagctgtta	gacaccctta	ttctgacgaa	gtagatagaa	gtattcgagt	ggtaagtcct	300
tgtggtatgt	gtaggaggtt	gatttcagac	tatgcaccag	attgttttgt	gttaatagaa	360
atgaatggca	agtttagtcaa	aactacgatt	gaagaactca	ttccactcaa	atatacccca	420
aattaaaagt	tttaccatac	caagcttggc	tgctgctga	ggctggacga	cctcgcgag	480
ttctaccggc	agtgcacatc	cgtcggcatc	caggaaacca	gcagcggcta	tcgcgcgcatc	540
catgcccccg	aactgcagga	gtggggaggg	acgatggccg	ctttgggtccg	gatcttttgtg	600
aaggaaacctt	acttctgtgg	tgtgacataa	ttggacaaac	tacctacaga	gatttaaagc	660
tctaaggtaa	atataaaaatt	tttaagtgtg	taatgtgtta	aactactgat	tctaatttgtt	720
tgtgtattttt	agatttccaa	ctatggaact	gatgaatggg	agcagtggtg	gaatgccttt	780
aatgaggaata	acctgttttg	ctcagaagaa	atgccatcta	gtgatgatga	ggctactgct	840
gactctcaac	attctactcc	tccaaaaaag	aagagaaaagg	tagaagaccc	caaggacttt	900
ccttcagaat	tgctaaagttt	tttgagtcatt	gctgtgttta	gtaatagaac	tcttgcttgc	960
tttgctattt	acaccacaaa	ggaaaaaagc	gcactgctat	acaagaaaat	tatggaaaaa	1020
tattctgtaa	cctttataag	taggcataac	agttataatc	ataacatact	gttttttctt	1080
actccacaca	ggcatagagt	gtctgctatt	aataactatg	ctcaaaaatt	gtgtactctt	1140
agctttttta	tttgtaaagg	ggtaataaag	gaattattga	tgtatagtgc	cttgactaga	1200
gatcataatc	agccatacca	catttgtaga	ggttttactt	gctttaaaaa	acctcccaga	1260
cctccccctg	aacctgaaac	ataaaaatgaa	tgcaatttgtt	gttggttaact	tgtttattgc	1320
agcttataat	ggttacaaat	aaagcaatag	catcacaaat	tcacacaaata	aagatccaga	1380
tttcgctcaa	gttagtataa	aaaagcaggg	ttcaatcctg	cagagaagct	tggcgccagc	1440
cggtctcaat	tgacacgggc	ccactagtga	gtcgtattac	gtagcttggc	gtaatcatgg	1500
tcatagctgt	ttcctgtgtg	aaattgttat	ccgctcacia	ttccacacaa	catcacagcc	1560
ggaagcataa	agtgtaaagc	ctgggtgtcc	taattgagtga	gctaactcac	attacatgtg	1620
agcaaaaaggc	cagcaaaaagg	ccaggaaccg	taaaaaggcc	gcgttgctgg	cgtttttcca	1680
taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	tcaagtccga	ggtggcgaaa	1740
cccagacaga	ctataaagat	accaggcggt	tccccctgga	agctccctcg	tgcgctctcc	1800
tgttccgacc	ctgccgctta	ccggatacct	gtccgccttt	ctcccttcgg	gaagcgtggc	1860
gcttttctcat	agctcacgct	gtaggtatct	cagttcgggtg	taggtcggtc	gctccaagct	1920
gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	gtaactatcg	1980
tcttgagtcc	aacccggtaa	gacacgactt	atcgccactg	gcagcagcca	ctggtaacag	2040
gattagcaga	gcgaggtatg	taggcggtgc	tacagagttc	ttgaagtgg	ggcctaacta	2100
cggctacact	agaagaacag	tatttgggtat	ctgcgctctg	ctgaagccag	ttaccttcgg	2160
aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	gtgggttttt	2220
tgtttgcaag	cagcagatta	cgcgagaaaa	aaaaggatct	caagaagatc	ctttgatctt	2280
ttctacgggg	tctgacgctc	agtggaaacga	aaactcacgt	taagggattt	tggtcatgag	2340
attatcaaaa	aggatcttca	cctagatcct	tttacgcgcc	ctgtagcggc	gcattaaagc	2400
cggcgggtgt	ggtggttacg	cgcagcgtga	ccgctacact	tgccagcgcc	ctagcgcccg	2460
ctcctttcgc	tttcttccct	tcctttctcg	ccacgttcgc	tttcccgcgc	aagctctaaa	2520
tcgggggctc	cctttaggggt	tccgatttag	tgctttacgg	cacctcgacc	ccaaaaaact	2580
tgatttgggt	gatggttcac	gtagtggggc	atcgccctga	tagacgggtt	ttcgcccttt	2640
gacgttggag	tccacgttct	ttaatagttg	actcttggtc	caaaactggaa	caaacctcaa	2700
ccctatctcg	ggctattctt	ttgatttata	agggattttg	ccgatttcgg	cctattgggt	2760
aaaaaatgag	ctgattttaac	aaaaatttaa	cgcgaatttt	aacaaaatat	taacgtttac	2820
aatttaaaata	tttgcttata	caatcttcct	gtttttgggg	cttttctgat	tatcaaccgg	2880
ggtaaatcaa	tctaaagtat	atatgagtaa	acttgggtctg	acagttacca	atgcttaatc	2940
agtgaggcac	ctatctcagc	gatctgtcta	tttcggttcat	ccatagttgc	ctgactcccc	3000
gtcgtgtaga	taactacgat	acgggagggc	ttaccatctg	gccccagtg	tgcaatgata	3060
ccgcgagacc	cacgctcacc	ggctccagat	ttatcagcaa	taaaccagcc	agccggaagg	3120
gccgagcgca	gaagtgggtc	tgcaacttta	tccgcctcca	tccagtcctat	taattgttgc	3180
cgggaagcta	gagtaagtag	ttcgccagtt	aatagtttgc	gcaacgttgt	tgccattgct	3240

-91-

```

acaggcatcg tgggtgtcacg ctcgtcgcttt ggtatggcctt cattcagctc cgggttcccaa 3300
cgatcaaggc gagttacatg atcccccatg ttgtgcaaaa aagcgggttag ctccttcggg 3360
cctccgatcg ttgtcagaag taagttggcc gcagtggttat cactcatggt tatggcagca 3420
ctgcataatt ctcttactgt catgccatcc gtaagatgct tttctgtgac tgggtgagtac 3480
tcaaccaagt cattctgaga atagtgtatg cggcgaccga gttgtctcttg cccggcgctca 3540
acacgggata ataccgcgcc acatagcaga actttaaaag tgctcatcat tggagaacgt 3600
tcttcggggc gaaaactctc aaggatctta ccgctgttga gatccagttc gatgtaaccc 3660
actcgtgcac ccaactgatc ttcagcatct tttactttca ccagcgtttc tgggtgagca 3720
aaaacaggaa ggcaaaatgc cgcaaaaaag ggaataaggg cgacacggaa atgttgaata 3780
ctcatactct tcctttttca atattattga agcatttatc agggttattg tctcatgagc 3840
ggatacatat ttgaatgtat ttagaaaaaa aaacaaatag gggttcccg cactttccc 3900
cgaaaagtgc cacctgacgt agttaacaaa ccaagccggg cctttattacc 3960
aagcgaagcg ccattcgcca ttcaggctgc gcaactgttg ggaaggcgga tccggtcggg 4020
cctcttcgct attacgccag ctggcgaaa ggggatgtgc tgcaaggcga ttaagttggg 4080
taacgccagg gttttcccag tcacgacgtt gtaaaacgac ggccagtcgc taatacgact 4140
cacttaaggc cttgactaga gggtcgacgc atgcctgtac atccggagac gcgtcacggc 4200
cgaagctagc gaattcgtgg atc 4223

```

<210> 122

<211> 2686

<212> DNA

<213> Artificial Sequence

<220>

<223> pUC18 Plasmid

<400> 122

```

tcgcgcgcttt cgggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca 60
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120
ttggcggttg tcggggctgg ccttaactatg gcgcatcaga gcagattgta ctgagagtgc 180
accatagtcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgcc 240
attcgccatt caggctgcgc aactgttggg aaggcgcatc ggtgcggggc tcttcgctat 300
tacgccagct ggcgaaaagg ggatgtgctg caaggcgatt aagttgggta acgccagggt 360
tttcccagtc acgacgttgt aaaacgacgg ccagtgcctaa gcttgcatgc ctgcaggctg 420
actctagagg atccccgggt accgagctcg aattcgtaat catggtcata gctgtttcct 480
gtgtgaaatt gttatccgct cacaattcca cacaacatac gagccggaag cataaagtgt 540
aaagcctggg gtgcctaatt agtgagctaa ctacatttaa ttgcgttgcg ctactgccc 600
gctttccagt cggaacacct gtcggtccag ctgcattaat gaatcgcca acgcgcgggg 660
agaggcggtt tgcgtattgg gcgctcttcc gcttcctcgc tcatgactc gctgcgctcg 720
gtcggttcggc tgcggcgagc ggtatcagct cactcaaaag cggtaatacg gttatccaca 780
gaatcagggg ataacgcagg aaagaacatg tgagcaaaag gccagcaaaa ggccagggaac 840
cgtaaaaagg ccgcgttgct ggctgttttc cataggtccc gccccctga cgagcatcac 900
aaaaatcgac gctcaagtca gaggtggcga aacccgacag gactataaag ataccaggcg 960
tttccccctg gaagctccct cgtgcgctct cctgttccga ccctgccgct taccggatac 1020
ctgtccgcct ttctcccttc ggggaagcgtg gcgctttctc atagctcacg ctgtagggtat 1080
ctcagttcgg tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc ccccgttcag 1140
cccgaccgct gcgccttacc cggtaactat cgtcttgagt ccaacccggg aagacacgac 1200
ttatcgccac tggcagcagc cactggtaac aggattagca gagcgaggta tgtaggcggt 1260
gctacagagt tcttgaaagt gtggcctaac tacggctaca ctagaaggac agtatttggt 1320
atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc 1380
aaacaaacca ccgctggtag cgggtggttt tttggttgca agcagcagat tacgcgcaga 1440
aaaaaaggat ctcaagaaga tcccttgatc ttttctacgg ggtctgacgc tcagtggaaac 1500
gaaaactcac gttaagggat tttggtcatg agattatcaa aaaggatctt cacctagatc 1560
cttttaaat aaaaatgaag ttttaaatca atctaaagta tatatgagta aacttggtct 1620
gacagttacc aatgcttaat cagttaggca cctatctcag cgatctgtct atttcgttca 1680
tccatagttg cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct 1740
ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca 1800
ataaacagc gcgcggaag ggccgagcgc agaaagtgtc ctgcaacttt atccgcctcc 1860
atccagctca ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtttg 1920
cgcaacgttg ttcgccattg tacaggcatc gttggtgcac gctcgtcggt tggtatggct 1980
tcattcagct cgggttccca acgatcaagg cagttacat gatccccc atgtgtgcaa 2040
aaagcggtta gctccttcgg tcctccgatc gttgtcagaa gtaagttggc cgcagtggtt 2100
tcactcatgg ttatggcagc actgcataat tctcttactg tcatgccaatc cgtaagatgc 2160
ttttctgtga ctggtgagta ctcaaccaag aatagtgat atccgcaccg gcggcgaccg 2220
agttgctctt gcccgcgctc aatacgggat aataccgcgc cacaatagcag aactttaaaa 2280
gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct accgctggtg accgctggtg 2340
agatccagtt cgatgtaacc cactcgtgca ccaactgat cttcagcatc ttttactttc 2400

```

-92-

accagcgttt	ctgggtgagc	aaaaacagga	aggcaaaatg	ccgcaaaaaa	gggaataagg	2460
gcgacacgga	aatgttgaat	actcatactc	ttcctttttc	aatattattg	aagcatttat	2520
caggggttatt	gtctcatgag	cggatacata	tttgaatgta	tttagaaaaa	taaacaaaata	2580
gggggttccgc	gcacatttcc	ccgaaaagtg	ccacctgacg	tctaagaaac	cattattatc	2640
atgacattaa	cctataaaaa	taggcgtatc	acgaggccct	ttcgtc		2686

<210> 123
<211> 8521
<212> DNA
<213> Artificial Sequence

<220>
<223> pCXeGFPattB(6xHS4)2 Plasmid

<400> 123

tacggggcg	gggatccact	agttattaat	agtaatcaat	tacgggggtca	ttagttcata	60
gcccatatat	ggagttccgc	gttacataac	ttacggtaaa	tggcccgccct	ggctgaccgc	120
ccaacgaccc	ccgccatttg	acgtcaataa	tgacgtatgt	tcccatagta	acgccaatag	180
ggactttcca	ttgacgtcaa	tgggtggact	atttacggta	aactgcccac	ttggcagtac	240
atcaagtgtg	tcatatgcca	agtacgcccc	ctattgacgt	caatgacggt	aaatggcccg	300
cctggcatta	tgcccagtac	atgaccttat	gggactttcc	tacttggcag	tacatctacg	360
tattagtcac	cgctattacc	atgggtcgag	gtgagcccca	cgttctgctt	cactctcccc	420
atctcccccc	cctccccacc	cccaattttg	tattttatta	ttttttaatt	attttgtgca	480
gcgatggggg	cggggggggg	ggggggcgcg	gccaggcggg	gcggggcggg	gcgagggggc	540
ggcgggggcg	aggcgagag	gtgcgggcg	agccaatcag	agcgcgcgcg	tccgaaagtt	600
tcctttttatg	gcgaggcggc	ggcgggcgcg	gccctataaa	aagcgaagcg	cgcgggcggc	660
gggagtcgct	gcgttgccct	cgcccccgtc	cccgcctcgc	gccgcctcgc	gccgccccgc	720
ccgctctga	ctgaccgct	tactcccaca	gggtgagcg	cgggacggcc	cttctcctcc	780
gggctgtaat	tagcgcttgg	tttaatgacg	gctcgcttct	tttctgtggc	tgcgtgaaag	840
ccttaaaggg	ctccgggagg	gcccttttgt	cgggggggag	cggtcggggg	gggtgcgtgcg	900
tgtgtgtgtg	cgtggggagc	gcgcgctgcg	cccgcgctgc	cccgcggtgc	gtgagcgtgc	960
cgggcgcggc	gcggggcctt	gtgcgctccg	cgtgtgcgcg	aggggagcgc	ggccgggggc	1020
gggtgccccg	gggtgcgggg	ggctgcgagg	ggaacaaagg	ctgcgtgcgg	gggtgtgtgcg	1080
tgggggggtg	agcagggggt	gtgggcgcgg	cggtcgggct	gtaaccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cggcccggtc	tgggtgcgg	ggctccgtgc	ggggcggtgc	1200
gcgggggtcg	ccgtgccggg	cggggggttg	cggcaggttg	gggtgccggg	cgggggcggg	1260
cgccctcggg	ccggggaggg	ctcgggggag	gggcgcggcg	gccccggagc	gcgggcggct	1320
gtcgaggcgc	ggcgagccgc	agccattgcc	ttttatggtg	atcgtgcgag	agggcgcgag	1380
gacttctctt	gtcccaaact	tggcggagcc	gaaatctggg	aggcgccgcg	gcacccccct	1440
tagcggcgcg	ggcggaagcg	gtgcggcgcc	ggcaggaagg	aaatgggcgg	ggagggcctt	1500
cgtgcgtcgc	cgcgccgcgc	tccccttctc	catctccagc	ctcggggctg	ccgcaggggg	1560
acggctgcct	tccggggggg	cggggcaggg	cggggttcgg	cttctggcgt	gtgaccggcg	1620
gctctagagc	ctctgtaaac	catgttcatg	cttcttctct	tttctacag	ctcctgggca	1680
acgtgctggt	tgttgtgctg	tctcatcatt	ttggcaaaga	attcgccacc	atggtgagca	1740
agggcgagga	gctgttcacc	ggggtggtgc	ccatcctggt	cgagctggac	ggcgacgtaa	1800
acggccacaa	gttcagcgtg	tccggcgagg	gcgaggcgga	tgccacctac	ggcaagctga	1860
ccctgaagtt	catctgcacc	accggcaagc	tggccgtgcc	ctggcccacc	ctcgtgacca	1920
ccctgacctg	cggcgtgcag	tgtctcagcc	gctaccccca	ccacatgaag	cagcacgact	1980
tcttcaagtc	cgccatgccc	gaaggctacg	tccaggagcg	caccatcttc	ttcaaggacg	2040
acggcaacta	caagaccgcg	gccgaggtga	agttcgaggg	cgacaccctg	gtgaaccgca	2100
tcgagctgaa	gggcatcgac	ttcaaggagg	acggcaacat	cctggggcac	aagctggagt	2160
acaactacaa	cagccacaac	gtctatatca	tggccgacaa	gcagaagaac	ggcatcaagg	2220
tgaacttcaa	gatccgccac	aacatcgagg	acggcagcgt	gcagctcgcc	gaccactacc	2280
agcagaacac	ccccatcgcc	gacggccccg	tgtgctgcc	cgacaaccac	tacctgagca	2340
ccagtcgcgc	cttgagcaaa	gaccccaacg	agaagcgcca	tcacatgggt	ctgctggagt	2400
tcgtgaccgc	cgccggggtc	actctcgcca	tggacgagct	gtacaagtaa	gaattcactc	2460
ctcaggtgca	ggctgcctat	cagaaggttg	tggctgggtg	ggccaatgcc	ctggctcaca	2520
aataccactg	agatcttttt	ccctctgcca	aaaatttatg	ggacatcatg	aagccccttg	2580
agcatctgac	ttctggctaa	taaaggaaat	ttattttcat	tgcaatagtg	tgttgggaatt	2640
ttttgtgtct	ctcactcgga	aggacatatg	ggagggcaaa	tcatttataaa	catcagaatg	2700
agtatttggg	ttagagtttg	gcaacatatg	ccatatgctg	gctgccatga	acaaaggtgg	2760
ctataaagag	gtcatcagta	tatgaaacag	ccccctgctg	tccattcctt	attccataga	2820
aaagccttga	cattgaggtta	gatttttttt	atatatttgt	ttgtgttatt	ttttcttta	2880
acatccctaa	acattgtttt	ctagccagat	ctagccagat	ttttcctcct	ctcctgacta	2940
ctcccagtcg	tagctgtccc	tcttctctta	tgaagatccc	tcgacctgca	gccccagctt	3000
ggcgtaatac	tggtcatagc	tgtttctctg	gtgaaattgt	tatccgctca	caattccaca	3060
caacatacga	gccggaagca	taaagtgtaa	agcctggggg	gcctaagtgc	tgagctaact	3120

cacattaatt	gcgttgcgct	cactgccccg	tttccagtcg	ggaaacctgt	cgtgccagcg	3180
gatccgcac	tcaattagtc	agcaaccata	gtcccccccc	taactccgcc	catccccgcc	3240
ctaactccgc	ccagttccgc	ccatttctccg	ccccatggct	gactaatttt	ttttattttat	3300
gcagaggccg	aggccgcctc	ggcctctgag	ctattccaga	agtagtgagg	aggctttttt	3360
ggaggctagt	ggatcccccg	ccccgtatcc	cccagggtgc	tgcaggctca	aagagcagcg	3420
agaagcggtc	agaggaaaagc	gatccccgtg	caccttcccc	gtgcccgggc	tgcccccgca	3480
cgctgccggc	tcggggatgc	ggggggagcg	ccggaccgga	gcggagcccc	gggaggctcg	3540
ctgctgcccc	ctagcggggg	agggacgtaa	ttacatccct	gggggctttg	ggggggggct	3600
gtccccgtga	gcggatccgc	ggccccgtat	ccccagggtg	tctgcaggct	caaagagcag	3660
cgagaagcgt	tcagaggaaa	gcgatccccg	gccaccttcc	ccgtgccccg	gctgtccccg	3720
cacgctgccg	gctcggggat	gcggggggag	cgccggaccg	gagcggagcc	ccggggcggt	3780
cgctgctgcc	ccctagcggg	ggagggacgt	aattacatcc	ctggggggctt	tggggggggg	3840
ctgtccccgt	gagcggatcc	gcggccccgt	atccccagg	tgtctgcagg	ctcaaagagc	3900
agcgagaaag	gttcagaggga	aagcgtatccc	gtgccacctt	ccccgtgccc	gggctgtccc	3960
cgcacgctgc	cggtcggggg	atgcgggggg	agcgccggac	cgagcggag	ccccggcg	4020
ctcgctgctg	ccccctagcg	ggggaggggac	gtaattacat	ccctgggggc	tttggggggg	4080
ggctgtcccc	gtgagcggat	ccgcggcccc	gtatccccca	gggtgtctgca	ggctcaaaga	4140
gcagcgagaa	cggttcagag	gaaagcgtatc	ccgtgccacc	ttccccgtgc	ccgggtgtgc	4200
cccgacacgt	gccgggtcgg	ggatgcgggg	ggagcgccgg	accggagcgg	agccccgggc	4260
ggctcgctgc	tgccccctag	cgggggagggg	acgtaatatc	atccctgggg	gctttggggg	4320
ggggctgtcc	ccctgagcgg	atccgcggcc	ccgtatcccc	cagggtgtctg	caggctcaaa	4380
gagcagcgag	aagcgttcag	aggaaagcga	tccccgtgcca	ccttccccgt	gccccgggctg	4440
tccccgcacg	ctgcccggctc	ggggatgcgg	ggggagcgcc	ggaccggagc	ggagccccgg	4500
gcggctcgct	gctgccccct	agcgggggag	ggacgtaat	acatccctgg	gggctttggg	4560
ggggggctgt	ccccgtgagc	ggatcccgcg	ccccgtatcc	cccagggtgtc	tgcaggctca	4620
aagagcagcg	agaagcgttc	agaggaaaagc	gatcccgtgc	caccttcccc	gtgcccgggc	4680
tgtccccgca	cgctgcgggc	tcggggatgc	ggggggagcg	ccggaccgga	gcggagcccc	4740
gggcccgtcg	ctgctgcccc	ctagcggggg	agggacgtaa	ttacatccct	gggggctttg	4800
ggggggggct	gtccccgtga	gcggatccgc	ggggctgcag	gaattcgatt	gaagcctgct	4860
tttttatact	aacttgagcg	aaatcaagct	cctaggcttt	tgcaaaaagc	taacttgttt	4920
attgcagctt	ataatgggta	caaataaagc	aatagcatca	caaatttcac	aaataaagca	4980
tttttttcac	tgcatcttag	ttgtggtttg	tccaaactca	tcaatgtatc	ttatcatgtc	5040
tggatccgct	gcattaatga	atcgcccaac	gcgcggggag	aggcgggttg	cgtattgggc	5100
gctcttcggc	ttcctcgctc	actgactcgc	tgcgctcggt	cgttcggctg	cggcgagcgg	5160
latcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	atcaggggat	aacgcaggaa	5220
agaacatgtg	agcaaaaaggc	cagcaaaaag	ccaggaaccg	taaaaaggcc	gcgttgctgg	5280
cgttttttcca	taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	tcaagtacga	5340
ggtggcgaaa	cccagacagga	ctataaagat	accaggcggt	tccccctgga	agctccctcg	5400
tgcgctctcc	tggttccgacc	ctgcccgtta	ccggatacct	gtccgccttt	ctccccctg	5460
gaagcgtggc	gctttctcaa	tgctcacgct	ttaggtatct	cagttcgggtg	taggtcggtc	5520
gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	5580
gtaactatcg	tcttgagtcc	aaccgggtaa	gacacgactt	atcgccactg	gcagcagcca	5640
ctggtaacag	gattagcaga	gcgagggtatg	taggcgggtg	tacagagttc	ttgaagtggt	5700
ggcctaacta	cggtacact	agaaggacag	tatttggtat	ctgcgctctg	ctgaagccag	5760
ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	5820
gtggtttttt	tggttgcaag	cagcagatta	cgcgagaaaa	aaaaggatct	caagaagatc	5880
tttaaatcaat	ttctacgggg	tctgacgctc	agtggaaacga	aaactcacgt	taagggattt	5940
ctttgatctt	attatcaaaa	aggatcttca	cctagatcct	tttaaatata	aatgaagtt	6000
tggatcatgag	ctaaaagtata	tatgagtaaa	ccttggtctga	cagttaccaa	tgcttaatca	6060
gtgaggcacc	tatctcagcg	atctgtctat	ttcggttcac	catagttgcc	tgactccccg	6120
tcgtgtagat	aactacgata	cgggagggct	taccatctgg	ccccagtgtc	gcaatgatac	6180
cgcgagaccc	acgctcaccc	gctccagatt	tatcagcaat	aaaccagcca	gccggaaggg	6240
ccgagcgag	aagtggctct	gcaactttat	ccgcctccat	ccagtctatt	aattgttgcc	6300
gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	caacgttggt	gccattgcta	6360
caggcatcgt	gggtgcacgc	tcgtcgtttg	gtatggcttc	attcagctcc	ggttcccaac	6420
gatcaaggcg	agttacatga	tcccccatgt	tgtgcaaaaa	agcggtagc	tccttcggtc	6480
ctccgatcgt	tgctcagaagt	aagttggccg	cagtggtatc	actcatgggt	atggcagcac	6540
tgcataatct	tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	ggtgagtagc	6600
caaccaagtc	attctgagaa	tagtgatgct	ggcgaccgag	ttgctcttgc	ccggcgctca	6660
tacgggataa	taccgcgcga	catagcagaa	ctttaaaagt	gctcatcatt	ggaaaacgtt	6720
cttcggggcg	aaaactctca	aggatcttca	cgctgttgag	atccagttcg	atgtaaccca	6780
ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	6840
aaacaggaag	gcaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	tggtgaatac	6900
tcataactct	cctttttcaa	tattattgaa	gcatttatca	gggttattgt	ctcatgagcg	6960
gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	ggttccgcgc	acatttcccc	7020
gaaaagtgcc	acctggctga	cggtatcgat	aagcttgata	tcgaattcct	gcagccccgc	7080
ggatccgctc	acggggagacg	ccccccccc	aagccccag	ggatgtaatt	acgtccctcc	7140

-94-

```

cccgcctaggg  ggcagcagcg  agccgccccg  ggctccgctc  cggctccggcg  ctccccccgc  7200
atccccgagc  cggcagcgtg  cggggacagc  ccgggcacgg  ggaaggtggc  acgggatcgc  7260
tttctcttga  acgcttctcg  ctgctctttg  agcctgcaga  cacctggggg  atacgggggc  7320
gcggatccgc  tcacggggac  agcccccccc  caaagccccc  agggatgtaa  ttacgtccct  7380
cccccgctag  ggggcagcag  cgagccgccc  ggggctccgc  tccggtccgg  cgtccccccc  7440
gcatccccga  gccggcagcg  tgcggggaca  gcccgggcac  ggggaaggtg  gcacgggatc  7500
gctttctctt  gaacgcttct  cgctgctctt  tgagcctgca  gacacctggg  ggatacgggg  7560
ccgcggtatc  gctcacgggg  acagcccccc  cccaaagccc  ccagggatgt  aattacgtcc  7620
ctcccccgct  agggggcagc  agcgagccgc  ccggggctcc  gctccgggtc  ggcgtcccc  7680
ccgcatcccc  gagccggcag  cgtgccccga  cagccccggc  acggggaagg  tggcacggga  7740
tcgctttctt  ctgaacgctt  ctgctgctc  tttgagcctg  cagacacctg  ggggatacgg  7800
ggccgcggt  ccgctcacgg  ggacagcccc  ccccccaggc  cccagggat  gtaattacgt  7860
ccctcccccg  ctagggggca  gcagcgagcc  gcccggggct  ccgctccggt  ccggcgctcc  7920
ccccgcctcc  ccgagccggc  agcgtgcccc  gacagcccg  gcacggggaa  ggtggcacgg  7980
gatcgctttc  ctctgaacgc  ttctcgctgc  tctttgagcc  tgcagacacc  tgggggatac  8040
ggggccgcg  atccgctcac  ggggacagcc  ccccccaaaa  gcccccagg  atgtaattac  8100
gtccctcccc  cgctaggggg  cagcagcgag  ccgccccggg  ctccgctccg  gtccggcgct  8160
ccccccgcat  ccccgagccg  gcagcgtgg  gggacagccc  gggcacgggg  aaggtggcac  8220
gggatcgctt  tctctgaac  gcttctcgct  gctctttgag  cctgcagaca  cctgggggat  8280
acggggccgc  ggtaccgctc  acggggacag  ccccccccca  aagccccag  ggatgtaatt  8340
acgtccctcc  ccgctaggg  ggcagcagcg  agccgccccg  ggctccgctc  cggcccgcg  8400
ctccccccgc  atccccgagc  cggcagcgtg  ccgggcacgg  ggaaggtggc  cacctggggg  8460
acgggatcgc  tttctctga  acgcttctcg  ctgctctttg  agcctgcaga  8520
a  8521

```

<210> 124
 <211> 8851
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> p18EPOcDNA Plasmid

```

<400> 124
cagttgcccc  cccgggtcgcg  cagggcggaac  tccccccccc  accggtgctc  gccgatctcg  60
gtcatggcgc  gcccgaggcg  gtcccggaag  tctgtggaca  cgacctccga  ccactcggcg  120
tacagctcgt  ccaggccgcg  caccacaccc  caggccaggg  tggtgtccgg  caccacctgg  180
tccctggacc  cgctgatgaa  caggggtcac  tccgtcccgga  ccacaccggc  gaagtccgtc  240
tccacgaagt  cccgggagaa  cccgagccgg  tccgtccaga  actcgaccgc  tccggcgacc  300
tcgcgcgcg  tgagcacggg  aacggcactg  gtcaacttgg  ccatggatcc  agatttccgt  360
caagttagta  taaaaaagca  ggcttcaatc  ctgcagagaa  gcttgatatc  gaattcctgc  420
agcccccgcg  atccgctcac  ggggacagcc  ccccccaaaa  gcccccagg  atgtaattac  480
gtccctcccc  cgctaggggg  cagcagcgag  ccgccccggg  ctccgctccg  gtccggcgct  540
ccccccgcat  ccccgagccg  gcagcgtgcg  gggacagccc  gggcacgggg  aaggtggcac  600
gggatcgctt  tctctgaac  gcttctcgct  gctctttgag  cctgcagaca  cctgggggat  660
acggggccgc  ggatccgctc  acggggacag  ccccccccca  aagccccag  ggatgtaatt  720
acgtccctcc  ccgctaggg  ggcagcagcg  agccgccccg  ggctccgctc  cggcccgcg  780
ctccccccgc  atccccgagc  cggcagcgtg  ccgggcacgg  ggaaggtggc  cacctggggg  840
acgggatcgc  tttctctga  acgcttctcg  ctgctctttg  agcctgcaga  cactcggggg  900
atacggggcc  gccgatccgc  tcacggggac  agcccccccc  caaagccccc  agggatgtaa  960
ttacgtccct  cccccgctag  ggggcagcag  cgagccgccc  ggggctccgc  tccggtccgg  1020
cgctcccccc  gcacccccga  gccggcagcg  tgccgggaca  gcccgggcac  ggggaaggtg  1080
gcacgggatc  gctttctct  gaacgcttct  cgctgctctt  tgagcctgca  gacacctggg  1140
ggatacgggg  ccgcggtatc  gctcacgggg  acagcccccc  cccaaagccc  ccagggatgt  1200
aattacgtcc  ctcccccgct  agggggcagc  agcgagccgc  ccggggctcc  gctccgggtc  1260
ggcgtcccc  ccgcatcccc  gagccggcag  cgtgccccga  cagccccggc  acggggaagg  1320
tggcacggga  tcgctttct  ctgaacgctt  ctgctgctc  tttgagcctg  cagacacctg  1380
gggatacgg  gcccgcggt  ccgctcacgg  ggacagcccc  cccccaaaag  cccagggat  1440
gtaattacgt  ccctcccccg  ctagggggca  gcagcgagcc  gcccggggct  ccgctccggg  1500
ccggcgctcc  ccccgcatcc  ccgagccgg  ctctgaacgc  tcttctgagc  gcacggggaa  1560
ggtaggcagg  gatcgcttcc  atccgctcac  cgctaggggg  ccccgagccg  gggacagccc  1620
tgggggatcc  ggggcccggg  atccgctcac  cgctaggggg  ccccgagccg  gggacagccc  1680
atgtaattac  gtccctcccc  ccccgagccg  tctctgaac  gcttctcgct  gttattaata  1740
gtccggcgct  aaggtggcac  cctgggggat  acggggcggg  gagttccgcg  ttacataact  1800
cctgggggat  tagttcatag  cccatatatg  gagttccgcg  ttacataact  tacggtaaat  1860
ggcccgctcg  1920

```

gctgaccgcc caacgacccc cgcccattga cgtcaataat gacgtatggt cccatagtaa 2040
cgccaatagg gactttccat tgacgtcaat ggttgacta ttacggtaa actgccact 2100
tggcagtaca tcaagtgtat catatgccaa gtacgcccc tattgacgtc aatgacggt 2160
aatggcccg ctaggttat gcccagtaca tgacctatg ggactttcct acttggcagt 2220
acatctacgt attagtcac gctattacca tgggtcgagg tgagccccc gttctgctt 2280
actctcccca tctccccccc ctccccaccc ccaattttgt atttatttat tttttaatta 2340
ttttgtgcag cgatgggggg gggggggggg ggggcccgc ccaggcgggg cggggcggg 2400
cgagggggcg ggcggggcga ggcggagagg tgcggcgga gccaatcaga ggcggcgct 2460
ccgaaagtgt ccttttatgg cgagggcggg gcggcgggcg ccctataaaa agcgaagcg 2520
gcgggcgggc ggagtcgctg cgttgccctt gccccgtgc ccgctccgcg ccgctcgcg 2580
ccgcccgcgc cggctctgac tgaccgcgtt actcccacag gtgagcgggc gggacggccc 2640
ttctcctcgg ggctgtaatt agcgcttggg ttaattgacg ctcgtttctt tctgtgggt 2700
gcgtgaaagc cttaaagggg tccgggaggg ccctttgtgc gggggggagc ggctcgggg 2760
gtgcgtgcgt gtgtgtgtgc gtggggagcg ccgctgcgg ccgctgcgc cggcggtg 2820
tgagcgctgc gggcgcgcg cggtcccg gtcgctccgc gtgtgcgcga ggggagcg 2880
gcccggggcg gtgccccgc gtgcggggg gctgcgagg gaacaaaggc tgcgtgcgg 2940
gtgtgtgcgt ggggggggtg gcaggggggt tgggcccgc ggtcggtg taaccccc 3000
ctgcaagctt cctcccgagt tgctgagcg ggcggcgctt cgggtgcggg gctcgtgcg 3060
gggctggtgc cggggtcgc cgtgcccggc ggggggtggc ggcaggtggg ggtgcccgg 3120
ggggcggggg cgctcgggg cggggagggg tccggggagg ggcgcggcg ccccgagcg 3180
ccggcggtgc tgcagcgcg cgcagcgca gccattgcct tttatggtaa tgcgtgcgg 3240
gggcgaggg acttcctttg tcccaaatct ggccgagccg aaatctggga ggcgcggcg 3300
caccctctc agcggggcg ggcgaagcg tgcggcgcc atctccagc aatggcggg 3360
gagggcctt gtgcgtgc cgcccgccgt ccccttctcc ggggttcggc tctcctgtg 3420
cgagggggga cggctgcctt cggggggggg aatgtcctgc tctgtgtgg ctctcctgt 3480
tgacggcggt ctctagaatg ggggtgcacg tctgtggcg agggccaagg atactactg 3540
cctgtcgtc cctcctctg ggctcccg tacctcttgg atatcactg agcaggccgt 3600
acagcggagt gctgtgctg acactgcagc gaggtcggg aggcctgtt aggcctgtt 3660
tctatgcctg gaagagatg ggaagctgt ctgcggggc ccgtcagtg tccagatgc 4020
ccctgtctg gcagctgcat gtggataaag cagaaggaag gctgacact tccgcaact 4080
ggagcggcgt aacaatcact aaagctgaag ctgtacacag gggaggcctg ggtgcctat 4140
gtacaagtaa gaattcact ctggctcaca aataccact agatctttt tcttggttaa 4200
ggcaatcatg aagccccttg agcatctgac ttttctgta tctactcga agacatatg 4260
tgcaatagtg tgttgaatt agtatttgg ctataaagag cttagagtt attttttt 4320
tcaattaaaa catcagaatg acaaagtggt attccataga acatccctaa ctttttctt 4380
gctgccatga acaaagtggt attccataga acatccctaa ctttttctt 4440
tccattcctt ttgttctatt ctctgacta ctcccagtc gcatgcctgc aggtcgact 4500
ttttcctcct tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 4560
ttttcctcct tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 4620
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 4680
ttttcctcct tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 4740
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 4800
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 4860
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 4920
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 4980
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 5040
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 5100
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 5160
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 5220
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 5280
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 5340
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 5400
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 5460
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 5520
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 5580
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 5640
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 5700
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 5760
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 5820
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 5880
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 5940
tcgacctgca gcccagctt tgcctgcagg ctcaaagagc agcgagaagc 6000

atgcggggggg	agcgccggac	cggagcggag	ccccggggcgg	ctcgctgctg	ccccctagecg	6060
ggggaggggac	gtaattacat	ccctggggggc	tttgggggggg	ggctgtcccc	gtgagcggat	6120
ccgcgggggct	gcaggaattc	gtaatcatgg	tcatagctgt	ttcctgtgtg	aaattgttat	6180
ccgctcacaa	ttccacacaa	catacgagcc	ggaagcataa	agtgtaaagc	ctgggggtgcc	6240
taatgagtga	gctaactcac	attaattgcg	ttgcgctcac	tgcccgcctt	ccagtcggga	6300
aacctgtcgt	gccagctgca	ttaatgaatc	ggccaacgcg	cggggagagg	cgggtttgcgt	6360
attgggcgct	cttcgcgttc	ctcgctcact	gactcgctgc	gctcggctcg	tcggctgcgg	6420
cgagcggat	cagctcactc	aaaggcggta	atacggttat	ccacagaatc	aggggataac	6480
gcaggaagaa	acatgtgagc	aaaaggccag	caaaaggcca	ggaaccgtaa	aaaggcccg	6540
ttgctggcgt	ttttccatag	gctccgcccc	cctgacgagc	atcacaaaaa	tcgacgctca	6600
agtcagaggt	ggcgaaaacc	gacaggacta	taaagatacc	aggcgtttcc	ccctggaagc	6660
tcctcgttgc	gctctcctgt	tccgaccctg	ccgcttaccg	gatacctgtc	cgcctttctc	6720
ccctcgggaa	gcgtggcgct	ttctcatagc	tcacgctgta	ggtatctcag	ttcgggtgtag	6780
gtcgttcgct	ccaagctggg	ctgtgtgcac	gaaccccccg	ttcagcccca	ccgctgcgcc	6840
ttatccggtc	gctctcgtct	tgagtccaac	ccggttaagac	acgacttatc	gccactggca	6900
gcagccactg	gtaacaggat	tagcagagcg	aggtatgtag	gcggtgctac	agagtctctg	6960
aagtgggtggc	ctaactacgg	ctacactaga	aggacagtat	ttgggtatctg	cgctctgctg	7020
aagccagtta	cccttcggaaa	aagagtgggt	agctcttgat	ccggcaaaaca	aaccaccgct	7080
ggtagcgggtg	gtttttttgt	ttgcaagcag	cagattacgc	gcagaaaaaa	aggatctcaa	7140
gaagatcctt	tgatcttttc	tacggggtct	gagcgtcagt	ggaacgaaaa	ctcacgttaa	7200
gggatttttg	tcgatgagatt	atcaaaaagg	atcttcacct	agatcctttt	aaatttaaaa	7260
tgaagtttta	aatcaatcta	aagtatatat	gagtaaactt	ggtctgacag	ttaccaatgc	7320
ttaatcagtg	agggacctat	ctcagcgatc	tgtctatttc	gttcatccat	agttgcctga	7380
ttccccgtcg	tgtagataac	tacgatacgg	gagggcttac	catctggccc	cagtgctgca	7440
atgataccgc	gagacccacg	ctcaccggct	ccagatttat	cagcaataaa	ccagccagcc	7500
ggaagggggcg	agcgcgagaag	tggtcctgca	actttatccg	cctccatcca	gtctattaat	7560
tggttcgggg	aagctagagt	aagtagttcg	ccagtttaata	gtttgcgcaa	cgttgtttgc	7620
attgctacag	gcctcgtggg	gtcacgctcg	tcggttggtg	tggtctcatt	cagctccggg	7680
tcccaacgat	caaggcgagt	tacatgatcc	ccctatgtgt	gcaaaaaagc	ggttagctcc	7740
ttcggctcctc	caatcgttgt	cagaagtaag	ttggccgcag	tggtatcact	catggttatg	7800
gcagcactgc	ataattctct	tactgtcatg	ccatccgtaa	gatgcttttc	tgtgactggg	7860
gagtaactcaa	ccaagtcat	ctgagaatag	tgtatgcggc	gaccgagttg	ctcttgcccg	7920
gcgtcaatac	gggataatac	cgcgccacat	agcagaactt	taaaagtgc	catcatgga	7980
aaacggttctt	cggggcgaaa	actctcaagg	atcttaccgc	tggtgagatc	cagttcgatg	8040
taaccacactc	gtgcacccaa	ctgatcttca	gcactcttta	ctttcaccag	cgtttctggg	8100
tgagcaaaaa	caggaaggca	aaatgccgca	aaaaagggaa	taagggcgac	acggaaaatg	8160
tgaatactca	tactcttcct	ttttcaatat	tattgaagca	tttatcaggg	ttattgtctc	8220
atgagcggat	acatatttga	atgtattttag	aaaaataaac	aaataggggt	tccgcgcaca	8280
ttccccgaa	aagtgccacc	tgacgtagtg	aacaaaaaaa	agcccgccga	agcgggcttt	8340
attaccaagc	gaagcgccat	tcgccattca	ggctgcgcaa	ctgttgggaa	gggctgacgg	8400
tgccgggcctc	ttcgctatta	cgccagctgg	cgaaaggggg	atgtgctgca	aggcgattaa	8460
gttgggtaac	gccaggggtt	tcccagtcac	gacgttgtaa	aacgacggcc	agtccgtaat	8520
acgactcact	taaggccctg	actagagggg	cgacggtata	cagacatgat	aagatacatt	8580
gatgagtttg	gacaaaccac	aactagaatg	cagtgaaaaa	aatgctttat	ttgtgaaatt	8640
tgtgatgcta	ttgcttttatt	tgtaacattt	ataagctgca	ataaacaagt	tgggggtggg	8700
gaagaactcc	agcatgagat	ccccgcgctg	gaggatcatc	cagccggcgt	cccggaaaaa	8760
gattccgaag	cccaaccttt	catagaaggc	ggcgtggaa	tcgaaatctc	gtagcacgtg	8820
tcagtcctgc	tcctcgccca	cgaagtgcac	g			8851

<210> 125

<211> 10474

<212> DNA

<213> Artificial Sequence

<220>

<223> p18genEPO Plasmid

<400> 125

cagttgcccgg	ccggggtcgcg	cagggcggaac	tccccgcccc	acgggtgctc	gccgatctcg	60
gtcatggccgg	gccccggaggc	gtccccgaag	ttcgtggaca	cgacctccga	ccactcggcg	120
tacagctcgt	ccaggcccgcg	cacccacacc	caggccaggg	tgttgtccgg	caccacctgg	180
tcctggaccg	cgctgatgaa	cagggtcacg	tcgtccccga	ccacaccggc	gaagtgcgtc	240
tccacgaagt	cccgggagaa	cccagcccg	tcggtccaga	actcgaccgc	tccggcgacg	300
tcgcgcgcgg	tgagcaccgg	aacggcactg	gtcaacttgg	ccatggatcc	agatttcgct	360
caagttagta	taaaaaagca	ggcttcaatc	ctgcagagaa	gcttgatc	gaattcctgc	420
agccccgcgg	atccgctcac	ggggacagcc	cccccccaaa	gccccagggg	atgtaattac	480
gtccctcccc	cgctaggggg	cagcagcgag	ccgcccgggg	ctccgctccg	gtccggcgct	540

-97-

ccccccgcat	ccccgagccg	gcagcggtgcg	gggacagccc	gggcacggggg	aaggtggcac	600
gggatcgctt	tccctctgaac	gcttctctcgt	gctcttttgag	cctgcagaca	cctgggggat	660
acggggccgc	ggatccgctc	acgggggacag	cccccccca	aagccccag	ggatgtaatt	720
acgtccctcc	cccgcctagg	ggcagcagcg	agccgcccgg	ggctccgctc	cggtccggcg	780
ctccccccgc	atccccgagc	cggcagcggtg	cggggacagc	ccgggcacgg	ggaaggtggc	840
acgggatccg	tttccctctga	acgcttctcg	ctgctctttg	agcctgcaga	cacctggggg	900
atacggggcc	gcggatccgc	tcacggggac	agcccccccc	caaagccccc	agggatgtaa	960
ttacgtccct	ccccgcctag	ggggcagcag	cgagccgccc	ggggctccgc	tccgggtccg	1020
cgctcccccc	gcatccccga	gcgggcagcg	tgccggggaca	gcccggggcac	ggggaggggtg	1080
gcacgggatc	gctttccctct	gaacgcttct	cgctgctctt	tgagcctgca	gacacctggg	1140
ggatacgggg	ccgcgggatcc	gctcacgggg	acagcccccc	cccaaagccc	ccaggggatgt	1200
aattacgtcc	ctcccccgct	agggggcagc	agcgagccgc	ccggggctcc	gctccgggtcc	1260
ggcgctcccc	ccgcatcccc	gagccggcag	cgtgcggggg	cagcccgggc	acgggggaagg	1320
tggcacggga	tcgcttttct	ctgaacgctt	ctcgctgctc	tttgagcctg	cagacacctg	1380
gggggacggg	ggccgcggat	ccgctcacgg	ggacagcccc	cccccaaagc	ccccagggat	1440
gtaattacgt	ccctcccccg	ctagggggca	gcagcgagcc	gcccggggct	ccgctccggt	1500
ccggcgctcc	ccccgcctcc	ccgagccggc	agcgtgcggg	gacagcccg	gcacggggaa	1560
gggtggcacgg	gatcgctttc	ctctgaacgc	ttctcgctgc	tctttgagcc	tgagagacc	1620
tgggggatac	ggggccgcgg	atccgctcac	ggggacagcc	cccccccaa	gccccaggg	1680
atgtaattac	gtccctcccc	cgctaggggg	cagcagcgag	ccgcccgggg	ctccgctccg	1740
gtccggcgct	cccccgcat	ccccgagccg	gcagcggtgc	gggacagccc	gggcacgggg	1800
aaggtggcac	gggatcgctt	tccctctgaac	gcttctcgt	gctctttgag	cctgcagaca	1860
cctgggggat	acggggcggg	ggatccacta	gttattaata	gtaatcaatt	acgggggtcat	1920
tagttcatag	cccatatatg	gagttccgcg	ttacataact	tacggtaaat	ggcccgctcg	1980
gctgaccgcc	caacgacccc	gcgccattga	cgtcaataat	gacgtatgtt	cccatagtaa	2040
cgccaatagg	gactttccat	tgacgtcaat	gggtggacta	tttacggtaa	actgccact	2100
tggcagtaca	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggt	2160
aatggccccg	ctggcattat	gcccagtaca	tgaccttatg	ggactttcct	acttggcagt	2220
acatctacgt	attagtcate	gctattacca	tgggtcgagg	tgagccccac	gttctgcttc	2280
actctcccc	tctccccccc	ctccccacc	ccaattttgt	atttatttat	ttttaatta	2340
ttttgtgcag	cgatgggggc	gggggggggg	ggggcgcgcg	ccaggcgggg	cggggcgggg	2400
cgagggggcg	ggcgggggga	ggcgggagag	tgccggcgga	gccaatcaga	gcggcgcgct	2460
ccgaaagtct	ccttttatgg	cgaggcgggc	gcggcgggcg	ccctataaaa	agcgaagcgc	2520
gcggcgggcg	ggagtcgctg	cggtgccttc	gccccgtgcc	ccgctccgcg	ccgcctcgcg	2580
ccgcccgcgc	cggtcttgac	tgaccgcgtt	actcccacag	gtgagcgggc	gggacggccc	2640
ttctctcccg	ggctgtaatt	agcgcttggg	ttaatgacgg	ctcgtttctt	ttctgtggct	2700
gcgtgaaagc	cttaaagggg	tccgggaggg	ccctttgtgc	ggggggggagc	ggctcggggg	2760
gtgcgtgcgt	gtgtgtgtgc	gtggggagcg	ccgcgtgcgg	ccgcgctgc	ccggcggtcg	2820
tgagcgctgc	gggcgcggcg	cggggctttg	tgcgctccgc	gtgtgcgcga	ggggagcgcg	2880
gcccggggcg	gtgcggggcg	gtgcgggggg	gctgcgaggg	gaacaaaggg	tgctgcgggg	2940
gtgtgtgcgt	gggggggtga	gcaggggggtg	tgggcgcggc	ggtcgggctg	taaccccccc	3000
ctgcaccccc	ctccccgagt	tgctgagcac	ggcccggctt	cgggtgcggg	gctccgtgcg	3060
ggggcggtgc	cggggctcgc	cgtgccgggg	gggggggtggc	ggcaggtggg	gggtgcgggg	3120
ggggcggggc	cgcctcgggc	cggggagggc	tccggggagg	ggcgcgggcg	ccccggagcg	3180
ccggcggtcg	tcgaggcgcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcgaga	3240
ggggcgaggg	acttcccttg	tcccaaatct	ggcgggagccg	aaatctggga	ggcgccggcg	3300
cacccccctc	agcgggcgcg	ggcgaaagcg	tgccggcgccg	gcaggaagga	aatgggcggg	3360
gagggccttc	gtgcgtcgcc	gcgcgcgct	ccccctctcc	atctccagcc	tcgggggtcg	3420
cgcaggggga	cggtcgctt	cggggggggc	ggggcagggc	gggggtcggc	ttctggcggt	3480
tgaccggcg	ctctagatgc	atgctcgagc	ggccgcaggt	gtgatggata	tctgcagaat	3540
tcgccccttc	tagaatgggg	gtgcacgggtg	agtactcgcg	ggctggggcg	ccccggccgc	3600
ccgggtccct	gtttgagcgg	ggatttagcg	ccccggctat	tggccaggag	gtggctgggt	3660
tcaaggaccg	gcgacttgct	aaggacccc	gaagggggag	gggggtgggg	tgccctccacg	3720
tgccagcggg	gacttggggg	agtccttggg	gatggcaaaa	acctgacctg	tgaaggggac	3780
acagtttggg	ggttgagggg	aagaagggtt	gggggttctg	ctgtgccagt	ggagaggaag	3840
ctgataagct	gataacctgg	gcgctggagc	caccacttat	ctgccagagg	ggaagcctct	3900
gtcacaccag	gattgaagtt	tggccggaga	agtggtatgt	ggtagctggg	gggtgggggt	3960
gcacacggga	gcaggattga	atgaaggcca	gggaggcagc	acctgagtcg	ttgcatgggt	4020
ggggacagga	aggacgagct	ggggcagaga	cgtggggatg	aagggaagctg	tccctccaca	4080
gccacccttc	tccctccccg	cctgactctc	agcctggcta	tctgttctag	aatgtcctgc	4140
ctggctgtgg	cttctcctgt	cctgctgtc	gctccctctg	ggcctcccg	tccgtggcgc	4200
cccaccacgc	ctcatctgtg	acagccgagt	cctggagagg	tacctcttgg	aggccaagga	4260
ggccgagaat	atcacgggtg	gaccccttcc	ccagcacatt	ccacagaact	cacgctcagg	4320
gcttcaggga	actcctccca	gatccaggaa	gggttggggg	gggttggggg	ggagtggga	4380
agctagacac	tgcccccccta	cataagaata	agtctgggtg	ccccaaacca	tacctggaaa	4440
ctaggcaagg	agcaaagcca	gcagatccta	cggcctgtgg	gccagggcca	gagccttcag	4500
ggacccttga	ctccccgggc	tgtgtgcatt	tcagacgggc	tgtgctgaac	actgcagctt	4560

-98-

gaatgagaat	atcactgtcc	cagacaccaa	agttaatttc	tatgcctgga	agaggatgga	4620
ggtgagttcc	tttttttttt	tttttctttt	cttttgga	atctcatttg	cgagcctgat	4680
tttggaatgaa	agggagaatg	atcgagggaa	aggtaaaatg	gagcagcaga	gatgaggctg	4740
cctgggcgca	gagggtcacg	tctataatcc	cagggtgaga	tggccgagat	gggagaattg	4800
cttgagccct	ggagtttcag	accaacctag	gcagcatagt	gagatcccc	atctctacaa	4860
acatttaaaa	aaattagtca	ggtgaagtgg	tgcatggtgg	tagtcccaga	tatttggaag	4920
gctgaggcgg	gaggatcgct	tgagcccagg	aatttgaggc	tgcatgagc	tgtgatcaca	4980
ccactgcact	ccagcctcag	tgacagagt	aggccctgtc	tcaaaaaaga	aaagaaaaaa	5040
gaaaaataat	gagggtgta	tggaatacat	tcattattca	ttcactcact	cactcactca	5100
ttcattcatt	cattcattca	acaagtctta	ttgcatacct	tctgtttgct	cagcttggtg	5160
cttggggctg	ctgaggggca	ggaggggag	ggtgacatgg	gtcagctgac	tcccagagtc	5220
cactccctgt	aggtcgggca	gcaggcctgg	gaagtcctgg	agggcctggc	cctgtctgctg	5280
gaagctgtcc	tgccggggcca	ggccctgttg	gtcaactcct	cccagccgtg	ggagcccctg	5340
cagctgcatg	tggataaagc	cgtcagtggc	cttcgcagcc	tcaccactct	gcttcgggct	5400
ctggggagccc	aggtgagtag	gagcgggacac	ttctgcttgc	cctttctgta	agaagggggag	5460
aagggctctg	ctaaggagta	caggaactgt	ccgtattcct	tccctttctg	tggcactgca	5520
gcgacctcct	gttttctcct	tggcagaagg	aagccatctc	ccctccagat	gcggcctcag	5580
ctgctccact	ccgaacaatc	actgctgaca	ctttccgcaa	actcttccga	gtcactcca	5640
atttccctcg	gggaaagctg	aagctgtaca	caggggaggc	ctgcaggaca	ggggacagat	5700
gacgtacaag	taagaattca	ctcctcaggt	gcaggctgcc	tatcagaagg	tgggtggctgg	5760
tgtggccaat	aggtcggtc	acaaatacca	ctgagatcct	tttccctctg	ccaaaaatta	5820
tggggacatc	atgaagcccc	ttgagcatct	gacttctggc	taataaagga	aattttatttt	5880
cattgcaata	gtgtgttggg	attttttgtg	tctctcactc	ggaaggacat	atgggagggc	5940
aaatcattta	aaacatcaga	atgagtattt	ggtttagagt	ttggcaacat	atgccatag	6000
ctggctgcca	tgaacaaagg	tggctataaa	gaggtcatca	gtatatgaaa	cagccccctg	6060
ctgtccattc	cttattccat	agaaaagcct	tgacttgagg	ttagattttt	tttatatttt	6120
gttttgtgtt	atttttttct	ttaacatccc	taaaattttc	cttacatggt	ttactagcca	6180
gatttttctc	cctctcctga	ctactcccag	tcatagctgt	ccctcttctc	ttatgaagat	6240
ccctcgacct	gcagcccaag	cttgcatgcc	tgcaggctga	ctctagtggg	tcccccgccc	6300
cgtatccccc	aggtgtctgc	aggctcaaag	agcagcgaga	agcgttcaga	ggaaagcgat	6360
cccgtgccac	cttccccgtg	cccgggctgt	ccccgcacgc	tgccggctcg	gggatgcggg	6420
gggagcgccg	gaccggagcg	gagccccggg	cggtctcgctg	ctgcccccta	gcgggggagg	6480
gacgtaatca	catcctggg	ggctttgggg	gggggctgtc	cccgtagcg	gatccgcggg	6540
cccgtatccc	ccaggtgtct	gcaggctcaa	agagcagcga	gaagcgttca	gaggaaagcg	6600
atcccgtgcc	accttccccg	tgcccgggct	gtccccgcac	gctgccgggt	cggggatgcg	6660
ggggggagcg	cggaccggag	cggagccccg	ggcggtctgc	tgctgcccc	tagcggggga	6720
gggacgtaat	tacatccctg	ggggctttgg	gggggggctg	tccccgtgag	cggatccgcg	6780
gccccgtatc	ccccagggtg	ctgcaggctc	aaagagcagc	gagaagcgtt	cagaggaaag	6840
cgatcccggt	ccaccttccc	cgtgcccgcc	ctgtccccgc	acgctgcccg	ctcgggggag	6900
cggggggagc	gccgggaccg	agcggagccc	cggggcgctc	gctgctgccc	cctagcgggg	6960
gagggagcgt	attacatccc	tgggggcttt	gggggggggc	tgtccccgtg	agcggatccg	7020
cggccccgtg	tccccagggt	gtctgcaggc	tcaaagagca	gcgagaagcg	ttcagaggaa	7080
agcgatcccg	tgccaccttc	cccggtcccc	ggctgtcccc	gcacgctgcc	ggctcggggg	7140
tgcggggggg	gcgcgggacc	ggagcggagc	cccgggcggc	tgcgtgctgc	cccctagcgg	7200
gggaggggag	taattacatc	cctgggggct	ttgggggggg	gctgtccccg	tgagcggatc	7260
cgcggccccg	tatccccag	gtgtctgcag	gctcaaagag	cagcgagaag	cgttcagagg	7320
aaagcgatcc	cgtgccacct	tccccgtgcc	cgggctgtcc	ccgcacgctg	ccggctcggg	7380
gatgcggggg	gagcgccgga	ccggagcggg	gccccgggcg	gctcgtgctg	gccccctagc	7440
gggggagggg	cgtaatatac	tccctggggg	ctttgggggg	gggctgtccc	cgtgagcggg	7500
tcccgggccc	cgtatcccc	aggtgtctgc	aggtcaaag	agcagcgaga	agcgttcaga	7560
ggaaagcgat	cccggtgccac	cttccccgtg	cccgggctgt	ccccgcacgc	tgccggctcg	7620
gggatgcggg	gggagcgccg	gaccggagcg	gagccccggg	cggctcgtctg	ctgcccccta	7680
gcgggggagg	gacgtaatta	catccctggg	ggctttgggg	gggggctgtc	cccgtagcgg	7740
gatccgcggg	gctgcaggaa	ttcgtaatca	tggtcatagc	tgtttcctgt	gtgaaattgt	7800
tatccgctca	caattccaca	caacatacga	gcgggaagca	taaagtgtaa	agcctggggg	7860
gcctaattgag	tgagctaact	cacattaatt	gcgttgcgct	cactgcccgc	tttccagtcg	7920
ggaaacctgt	cgtgccagct	gcattaatga	atcgcccaac	gcgcggggag	aggcggtttg	7980
cgtattgctg	ttcctccgc	ttcctcgctc	actgactcgc	tgcgctcggt	cgttcgggtg	8040
cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	atcaggggat	8100
aacgcaggaa	agaacatgtg	agcaaaaagg	cagcaaaaagg	ccaggaaccg	taaaaaggcc	8160
gcgttgctgg	agtttttcca	taggtccgc	ccccctgacg	agcatcaca	aaatcgacgc	8220
tcaagtcaga	ggtggcgaaa	cccagacagg	ctataaagat	accaggcgtt	tccccctgga	8280
agctccctcg	tgcgctctcc	tggtccgacc	ctgcccgtta	ccggatacct	gtccgccttt	8340
ctccctcgg	gaagcgtggc	gctttctcat	agctcacgct	gtaggatatc	cagttcgggtg	8400
taggtcgttc	gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	8460
gccttatccg	gtaactatcg	tcttgagtcg	aaccgggtaa	gacacgactt	atcgccactg	8520
gcagcagcca	ctggtaacag	gattagcaga	gcgaggtatg	taggcgggtgc	tacagagttc	8580

-99-

```

ttgaagtggg  ggccctaacta  cggctacact  agaaggacag  tatttggtat  ctgcgctctg  8640
ctgaagccag  ttaccttcgg  aaaaagagtt  ggtagctctt  gatccggcaa  acaaacacc  8700
gctggtagcg  gtgggttttt  tgtttgcaag  cagcagatta  cgccgagaaa  aaaaggatct  8760
caagaagatc  ctttgatctt  ttctacgggg  tctgacgctc  agtggaacga  aaactcacgt  8820
taagggattt  tgggtcatgag  attatcaaaa  aggatcttca  cctagatcct  tttaaattaa  8880
aatgaagtt  ttaaataaat  ctaaagtata  tatgagtaaa  cttgggtctga  cagttaccaa  8940
tgcttaatac  gtgaggcacc  tatctcagcg  atctgtctat  ttcgttcac  catagttgcc  9000
tgactccccg  tcgtgtagat  aactacgata  cgggagggct  taccatctgg  cccagtgct  9060
gcaatgatac  cgcgagaccc  acgctcaccg  gctccagatt  tatcagcaat  aaaccagcca  9120
gccggaaggg  ccgagcgcag  aagtggtcct  gcaactttat  ccgcctccat  ccagttctatt  9180
aattgttgcc  gggaagctag  agtaagtagt  tcgccagtta  atagtttgcg  caacgttggt  9240
gccattgcta  caggcatcgt  ggtgtcacgc  tcgtcgtttg  gtatggcttc  attcagctcc  9300
ggttcccaac  gatcaaggcg  agttacatga  tcccccatgt  tgtgcaaaaa  agcgggttagc  9360
tccttcggtc  ctccgatcgt  tgtcagaagt  aagttggccg  cagtgttatc  actcatgggt  9420
atggcagcac  tgcataattc  tcttactgtc  atgccatccg  taagatgctt  ttctgtgact  9480
ggtagtact  caaccaagtc  attctgagaa  tagtgtatgc  ggccgaccgag  ttgctcttgc  9540
ccggcgctca  tacgggataa  taccgcgcca  catagcagaa  ctttaaaagt  gctcatcatt  9600
ggaaaacgtt  cttcggggcg  aaaactctca  aggatcttac  cgctgttgag  atccagttcg  9660
atgtaaccca  ctctgtcacc  caactgatct  tcagcatctt  ttactttcac  cagcgtttct  9720
gggtgagcaa  aaacaggaag  gcaaaatgcc  gcaaaaaagg  gaataagggc  gacacggaaa  9780
tgttgaatac  tcataactct  cttttttcaa  tattattgaa  gcatttatca  gggttattgt  9840
ctcatgagcg  gatacatatt  tgaatgtatt  tagaaaaata  aacaaatagg  ggttccgcgc  9900
acatttcccc  gaaaagtgcc  acctgacgta  gttaacaaaa  aaaagcccg  cgaagcgggc  9960
tttattacca  agcgaagcgc  cattcgccat  tcaggcttgc  caactgttgg  gaagggcgat  10020
cgggtgcggg  ctcttcgcta  ttacgccagc  tggcgaaagg  gggatgtgct  gcaaggcgat  10080
taagttgggt  aacgccaggg  ttttcccgat  cagcagcttg  taaaacgacg  gccagtcctg  10140
aatacgactc  acttaaggcc  ttgactagag  ggtcgacggg  atacagacat  gataagatac  10200
attgatgagt  ttggacaaac  cacaactaga  atgcagtga  aaaaatgctt  tatttgtgaa  10260
atttgtgatg  ctattgcttt  atttghtaac  attataagct  gcaataaaca  agttgggggtg  10320
ggcgaagaac  tccagcatga  gatccccgcg  cttggaggatc  atccagccgg  cgtcccggaa  10380
aacgattccg  aagcccaacc  tttcatagaa  ggccgcggtg  gaatcgaaat  ctctgtagcac  10440
gtgtcagtc  tgctcctcgg  ccacgaagtg  caccg  10474

```

<210> 126

<211> 6119

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeoeGFP Plasmid

<400> 126

```

cagttgccgg  ccgggtcggg  cagggcgaac  tcccccccc  acggctgctc  gccgatctcg  60
gtcatggcgg  gcccgaggcg  gtcccggaag  ttctgtggca  cgacctccga  ccactcggcg  120
tacagctcgt  ccaggccgct  caccacacac  caggccaggg  tgttgtccgg  caccacctgg  180
tcctggaccg  cgtgatgaa  cagggtcacg  tcgtcccgga  ccacaccggc  gaagtcgtcc  240
tccacgaagt  cccgggagaa  cccgagccgg  tcggtccaga  actcgaccgc  tccggcgacg  300
tcgcgcgcgg  tgagcaccgg  aacggcactg  gtcaacttgg  ccattggatcc  agatttcgct  360
caagttagta  taaaaaagca  ggcttcaatc  ctgcagagaa  gcttgggctg  caggtcgagg  420
gatcttcata  agagaagagg  gacagctatt  actgggagta  gtcaggagag  gaggaaaaat  480
ctggctagta  aaacatgtaa  ggaaaatttt  agggatgtta  aagaaaaaaa  taacacaaaa  540
caaaatataa  aaaaaatcta  acctcaagtc  aaggcttttc  tatggaataa  ggaatggaca  600
gcagggggct  gtttcatata  ctgatgacct  ctttatagcc  acctttgttc  atggcagcca  660
gcatatggca  tatgttgcca  aactctaaac  caaataactca  ttctgatgtt  ttaaattgatt  720
tgccctccca  tatgtccttc  cgagtgaag  acacaaaaaa  ttccaacaca  ctattgcaat  780
gaaaataaat  ttcttttatt  agccagaagt  cagatgctca  aggggcttca  tgatgtcccc  840
ataatttttg  gcagagggaa  aaagatctca  gtggatattg  tgagccaggg  cattggccac  900
accagccacc  accttctgat  aggcagctcg  caactgagga  gtgaattott  acttgtacag  960
ctcgtccatg  ccgagagtga  tcccggcggc  ggtaacgaac  tccagcagga  ccatgtgatc  1020
cgccttctcg  ttggggctct  tgctcagggc  ggtctgggtg  ctcaggtagt  ggttgtcggg  1080
cgacagcacg  gggcggtcgc  cgatgggggt  gttctgctgg  tagtggtcgg  cgaggtgcac  1140
gctgccgtcc  tcgatgttgt  ggcggatctt  gaagttcacc  ttgatgccgt  tcttctgctt  1200
gtcggccatg  atatagacgt  tgtggctggt  gtatgtgtac  tccagcttgt  gccccaggat  1260
ggtgccgtcc  tccttgaagt  cgatgccctt  cagctcgatg  cgggtcacca  ggggtgtcgc  1320
ctcgaacttc  acctcggcgc  ggggtcttga  gttgccgtcg  tccttgaaga  agatgggtcg  1380
ctcctggacg  tagccttcgg  gcatggcgga  cttgaagaag  tcgtgctgct  tcatgtggct  1440
ggggtagcgg  ctgaagcact  gcaagccgta  ggtcagggtg  gtcacgaggg  tgggcccagg  1500

```


-100-

cacgggacgc	ttgcccgtgg	tgcagatgaa	cttcagggtc	agcttgccgt	aggtggcatc	1560
gcccctcgccc	tgcgcccga	cgctgaactt	gtggccggtt	acgtcgccgt	ccagctcgac	1620
caggatgggc	accaccccg	tgaacagctc	ctcgcccttg	ctcaccatgg	tggcgaattc	1680
tttgccaaaa	tgatgagaca	gcacaacaac	cagcacgttg	cccaggagct	gtaggaaaaa	1740
gaagaaggca	tgaacatgg	tagcagaggg	tctagagccg	ccggtcacac	gccagaagcc	1800
gaaccccgcc	ctgccccgtc	ccccccgaag	gcagccgtcc	ccctgcggca	gccccgaggc	1860
tggagatgga	gaaggggacg	gcgggcgccg	gacgcacgaa	ggccctcccc	gcccattttc	1920
ttcctgccc	cgccgcaccg	cttcgccccg	gcccgc taga	gggggtgcgg	cggcgccctc	1980
cagattttcgg	ctccgcccaga	tttggggacaa	aggaagtccc	tgcgcccctc	cgcacgatta	2040
ccataaaagg	caatggctgc	ggctcgccgc	gcctcgacag	ccgcccggcg	tccggggccg	2100
ccgcgcccct	cccccgagcc	ctccccggcc	cgaggcgccc	ccgccccgcc	cggcaccccc	2160
acctgcccgc	accccccgcc	cggcacggcg	agcccccgcc	cacgccccgc	acggagcccc	2220
gcacccgaag	ccggggccgtg	ctcagcaact	cgggggagggg	gggtgcagggg	gggggttacag	2280
cccgcacgcc	gcgcccacac	cccctgctca	ccccccacg	cacacacccc	gcacagcagc	2340
tttgttcccc	tgcgagcccc	cccgacccgc	ggggcacccg	ccccggccgc	gctccccctg	2400
cgcacacgcg	gagcgacaaa	agcccccgcc	cgcgcccgcg	gcgctcacag	ccgcccgggca	2460
gcgccccg	cacgcggcgc	tccccacgca	cacacacacg	cacgcacccc	ccgagccgct	2520
cccccccgca	caaagggccc	tcccgagacc	ctttaaggct	ttcacgcagc	cacagaaaag	2580
aaacgagccg	tcaataaacc	aagcgcta	tacagcccg	aggagaaggg	ccgtcccgcc	2640
cgctcacctg	tgggagtaac	gcggtcagtc	agagccgggg	cgggcggcgc	gaggcgccgc	2700
ggagcggggc	acggggcgaa	ggcaacgcag	cgactcccgc	ccgcccgcgc	cttcgctttt	2760
tataggcccg	ccgcccgcgc	cgctcgcca	taaaaggaaa	ctttcgggag	gcgcccgtct	2820
gattggctgc	cgccgcacct	ctccgcctcg	ccccgccccg	ccccctgccc	cgccccgccc	2880
cgctggggcg	gccccccccc	gccccccg	cccatcgctg	cacaaaaata	ttaaaaaata	2940
aataaaataca	aaattggggg	tggggagggg	ggggagatgg	ggagagtga	gcagaacgtg	3000
gggtcacctc	cgacccatgg	taatagcgat	gactaatacg	tagatgtact	gccaaagtgg	3060
aaagtcccat	aaggatcatgt	actgggcata	atgccaggcg	ggccatttac	cgctattgac	3120
gtcaataggg	ggcgactctg	gcataatgata	cacttgatgt	actgccaa	gggcagttta	3180
ccgtaaatag	tccacccatt	gacgtcaatg	gaaagtccct	attggcggtt	ctatgggaac	3240
atagctcatt	attgacgtca	atgggcgggg	gtcggtgggg	ggtcagccag	cggggccatt	3300
taccgtaagt	tatgtaacgc	ggaactccat	atatgggcta	tgaactaatg	accccgta	3360
tgattactat	taataactag	aggatccccg	ggatccgagc	togaattcgt	aatcatggtc	3420
atagctgttt	cctgtgtgaa	attgttatcc	gtccacaatt	ccacacaaca	tacgagccgg	3480
aagcataaag	tgtaaagcct	gggtgacct	atgagtgagc	taactcacat	taattgcgtt	3540
gcgctcactg	cccgttttcc	agtccggaaa	cctgtcgtgc	cagctgcatt	aatgaatcgg	3600
ccaacggcg	gggagagcg	gtttgcgtat	tgggcgctct	tccgcttcc	cgctcactga	3660
ctcgctgcgc	tccgtcgctt	ggctgcggcg	agcggtatca	gctcactcaa	aggcggtaat	3720
acggttatcc	acagaatcag	gggataacgc	aggaaagaac	atgtgagcaa	aaggccagca	3780
aaagggcagg	aacgtaaaaa	aggccgcgtt	gtcggcggtt	ttccataggc	tccgcccccc	3840
tgacgagcat	cacaaaaatc	gacgctcaag	tcagaggtgg	cgaaacccga	caggactata	3900
aagataccag	gcgtttcccc	ctggaaagct	cctcggtgcg	tctcctgttc	cgacccctgc	3960
gcttacccga	tacctgtccg	cttttctccc	ttcgggaagc	gtggcgcttt	ctcatagctc	4020
acgctgtagg	tatctcagtt	cggtgtaggt	cgttcgctcc	aagctgggct	gtgtgcacga	4080
acccccggtt	cagcccagacc	gctgcgcctt	atccgggta	tatcgtcttg	agtcacaccc	4140
ggtaagacac	gacttatcgc	cactggcagc	agccactggt	aacaggatta	gcagagcgag	4200
gtatgtaggc	ggtgctacag	agttcttgaa	gtgggtggct	aactacggct	acactagaag	4260
gacagtattt	ggatatctgc	ctctgctgaa	gccagttacc	ttcggaaaaa	gagtgggtag	4320
ctcttgatcc	ggcaaacaaa	ccaccgctgg	tacgggtgg	ttttttgttt	gcaagcagca	4380
gattacgcgc	agaaaaaaag	gatctcaaga	agatcctttg	atcttttcta	cggggtctga	4440
cgctcagtg	aacgaaaaact	cacgttaagg	gattttgggt	atgagattat	caaaaaggat	4500
cttcacctag	atccttttaa	attaaaaatg	aagtttttaa	tcaatctaaa	gtatatatga	4560
gtaaacttgg	tctgacagtt	accaatgctt	aatcagtgag	gcacctatct	cagcgatctg	4620
tctatttcgt	tcatccatag	ttgcctgact	ccccgtcggt	tagataacta	cgatacggga	4680
gggcttacca	tctggcccca	gtgctgcaat	gataccgcga	gacccacgct	caccggctcc	4740
agattttatca	gcaataaaacc	agccagccgg	aagggccgag	cgcagaagtg	gtcctgcaac	4800
tttatccggc	tccatccagt	ctattaattg	ttgcccggaa	gctagagtaa	gtagttcgcc	4860
agttaaatag	ttgcgcaacg	ttgttgccat	tgtcacaggc	atcgtgggtg	cacgctcgct	4920
gtttggatag	gcttcattca	gctccggttc	ccaacgatca	aggcgagtta	catgatcccc	4980
catgttgtgc	aaaaaaagcg	ttagctcctt	cggtcctccg	atcgttgtca	gaagtaagtt	5040
ggccgagtg	ttatcactca	tggttatggc	agcactgcat	aattctctta	ctgtcatgcc	5100
atccgtaaga	tgtttttctg	tgactgggtg	gtactcaacc	aagtcattct	gagaatagtg	5160
tatgcggcga	ccgagttgct	cttgcggcgc	gtcaatacgg	gataataccg	cgccacatag	5220
cagaacttta	aaagtgtcca	tcatgggaaa	acgttcttcg	gggcgaaaaa	tctcaaggat	5280
cttaccgctg	ttgagatcca	gttcgatgta	acccactcgt	gcacccaact	gatcttcagc	5340
atcttttact	ttcaccagcg	tttctgggtg	agcaaaaaa	ggaaggcaaa	atgccgcaaa	5400
aaagggaata	agggcgacac	ggaaatgttg	aatactcata	ctcttctctt	ttcaatatta	5460
ttgaagcatt	tatcagggtt	attgtctcat	gagcggatag	atatttgaat	gtatttagaa	5520

-101-

```

aaataaaca ataggggttc cgcgcacatt tccccgaaaa gtgccacctg acgtagtta 5580
caaaaaaaag cccgccgaag cgggctttat taccagcgga agcgccattc gccattcagg 5640
ctgcgcaact gttgggaagg gcgatcggtg cgggcctctt cgctattacg ccagctggcg 5700
aaagggggat gtgctgcaag gcgattaagt tgggtaacgc caggggtttc ccagtcacga 5760
cgttgtaaaa cgacggccag tccgtaatac gactcactta aggccttgac tagaggggtc 5820
acggtataca gacatgataa gatacattga tgagtttgga caaaccacaa ctagaatgca 5880
gtgaaaaaaa tgctttattt gtgaaatttg tgatgctatt gctttatttg taaccattat 5940
aagctgcaat aaacaagttg ggggtgggcga agaactccag catgagatcc ccgcgctgga 6000
ggatcatcca gccggcggtc cggaaaaaga ttccgaagcc caacctttca tagaaggcgg 6060
cgggtggaatc gaaatctcgt agcacgtgtc agtctctgctc ctcgccacag aagtgcacg 6119

```

<210> 127

<211> 5855

<212> DNA

<213> Artificial Sequence

<220>

<223> pCXIamInt Plasmid (Wildtype Integrase)

<400> 127

```

gtcgacattg attattgact agttattaat agtaatcaat tacgggggtca ttagttcata 60
gcccataat ggagttccgc gttacataac ttacggtaaa tggcccgccct ggctgaccgc 120
ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180
ggactttcca ttgacgtcaa tgggtgggact atttacggta aactgcccac ttggcagtac 240
atcaagtgtg tcatatgcca agtacgcccc ctattgacgt caatgacggg aaatggcccc 300
cctggcatta tgcccagtac atgaccttat gggactttcc tacttggcag tacatctacg 360
tattagtcac cgctattacc atgggtcgag gtgagcccca cgctctgctt cactctcccc 420
atctcccccc cctccccacc cccaattttg tattttatta ttttttaatt attttgtgca 480
gcgatggggg cggggggggg ggggggcgcg gccaggcggg gcgggggcgg gcgagggggc 540
gggcggggcg aggcgggagag gtgcggcggc agccaatcag agcggcgcgc tccgaaagt 600
tccttttatg gcgagggcgc ggcgggcggc gccctataaa aagcgaagcg cgcggcgggc 660
gggagtcgct gcgttgccct cgcgccgtgc ccgctccgc gccgcctcgc gccgcccgc 720
ccggtctga ctagcccgct tactcccaca ggtgagcggg cgggacggcc ctctctccc 780
gggctgtaat tagcgcttgg tttaatgacg gctcgtttct tttctgtggc tgcgtgaaag 840
ccttaaaggg ctccgggagg gccctttgtg cggggggggg cggtcggggg ggtgcgtgcg 900
tgtgtgtgtg cgtgggggagc gtcgcgtgcg gccgcgctg cccggcggtg gtgagcgctg 960
cgggcgcggc gcggggcctt gtgcgtccg cgtgtgcgag aggggagcgc ggccgggggc 1020
gggtgccccg ggtgcggggg ggctgcgagg ggaacaaagg ctgcgtgcgg ggtgtgtgcg 1080
tggggggggg agcggggggt gtggcgcgcg cggtcgggct gtaaccccc cctctccc 1140
cctccccgag ttgctgagca cggcccggtc tcgggtgcgg ggctccgtgc ggggcgtggc 1200
gcggggctcg ccgtgcgggg cgggggggtg cggcaggtgg gggtgcgggg cggggcgggg 1260
ccgctcggg ccggggaggg ctccggggcg gggcgcgggc gccccggagc gcccgcggt 1320
gtcgagggcg ggcgagcgcg agccattgcc ttttatggta atcgtgcgag agggcgcgag 1380
gacttccttt gtcccaaacc tggcgggagc gaaatctggg aggcgcggcc gcacccccct 1440
tagcgggagc gggcgaaagc gtgcggcgcc ggcaggaagg aaatgggcgg ggaaggccct 1500
cgtgcgtcgc cgcgcgcgcg tccccttctc catctccagc ctcggggctg ccgcaggggg 1560
acggctgcct tcggggggga cggggcgagg cgggggtcgg ctctcgccgt gtgacgggcg 1620
gctctagagc ctctgctaac catgttcatg ccttctctt tttctacag ctctgggca 1680
acgtgctggt tgtgtgtgct tctcatcatt ttggcaaaaga attcatggga agaaggcgaa 1740
gtcatgagcg ccgggattta ccccttaacc tttatatagg aaacaatgga tattactgct 1800
acaggagacc aaggacgggt aaagagtttg gattaggcag agacaggcga atcgcaatca 1860
ctgaagctat acaggccaac attgagttat ttccaggaca caaacacaag cctctgacag 1920
cgagaatcaa cagtgataat tccgttacgt tacattcatg gcttgatcgc tacgaaaaaa 1980
tccggccag cagaggaatc aagcagaaga cactcataaa ttacatgagc aaaattaaag 2040
caataaggag gggctctgct gatgctccac ttgaagacat caccacaaaa gaaattgcgg 2100
caatgctcaa tggatacata gacgagggca aggcggcgct agccaagtta atcagatcaa 2160
cactgagcga tgcattccga gaggcaatag ctgaaggcca tataacaaca aacctgtcgt 2220
ctgccactcg cgcagcaaaa tcagaggtaa ggagatcaag acttacgggt gacgaatacc 2280
tgaaaattta tcaagcagca gaatcatcac catgttggct cagacttgca atggaactgg 2340
ctgttgttac cgggcaacga gttggtgatt tatgcgaaat gaagtgggtc gatatcgtag 2400
atggatatct ttatgtcgag caaagcaaaa caggcgtaaa aattgccatc ccaacagcat 2460
tgcatattga tgctctcgga atatcaatga aggaacact tgataaatgc aaagagattc 2520
ttggcggaga aaccataatt gcatctactc gtcgcgaacc gctttcatcc ggcacagtat 2580
caaggatttt tatgcgcgca cgaaaagcat caggtctttc ctccgaaggg gatcccgcta 2640
cctttcacga gttgcgcagt ttgtctgcaa gactctatga gaagcagata agcgataagt 2700
ttgtcaaca tcttctcggg cataagtcgg acaccatggc atcacagtat cgtgatgaca 2760
gaggcagggg gtgggacaaa attgaaatca aataagaatt cactcctcag gtgcaggctg 2820

```

-102-

```

cctatcagaa ggtggtggct ggtgtggcca atgccctggc tcacaaatac cactgagatc 2880
ttttccctc tgccaaaaat tatggggaca tcatgaagcc ccttgagcat ctgacttctg 2940
gctaataaag gaaatatttatt ttcatgtgaa tagtgtgttg gaattttttg tgtctctcac 3000
tcggaaggac atatgggagg gcaaatcatt taaaacatca gaatgagtat ttgggtttaga 3060
gtttggcaac atatgccata tgctggctgc catgaacaaa ggtggctata aagagggtcat 3120
cagtatatga aacagccccc tgctgtccat tccttattcc atagaaaagc cttgacttga 3180
ggttagatbt tttttatatt ttgttttggtg ttattttttt ctttaacatc cctaaaaattt 3240
tccttacatg ttttactagc cagatttttc ctctctcct gactactccc agtcatagct 3300
gtccctcttc tcttatgaag atccctcgac ctgcagccca agcttggcgt aatcatgggtc 3360
atagctgttt cctgtgtgaa attgttatcc gctcacaatt ccacacaaca tacgagccgg 3420
aagcataaag tgtaaagcct ggggtgccta atgagtgcgc taactcacat taattgcgtt 3480
gcgtccactg cccgctttcc agtcgggaaa cctgtcgtgc cagcgggatcc gcatctcaat 3540
tagtcagcaa ccatagtccc gccctaact ccgcccattcc cgcccctaac tccgcccagt 3600
tcgcccatt ctccgcccc tggctgacta atttttttta tttatgcaga ggccgaggcc 3660
gcctcggcct ctgagctatt ccagaagtag tgaggaggct tttttggagg cctaggcttt 3720
tgcaaaaagc taacttgttt attgcagctt ataattgggtta caaataaagc aatagcatca 3780
caaatttcac aaataaagca tttttttcac tgcattctag ttgtggtttg tccaaactca 3840
tcaatgtatc ttatcatgtc tggatccgct gcattaatga atcgcccaac gcgcggggag 3900
aggcggtttg cgtattgggc gctcttccgc ttctctgctc actgactcgc tgcgtcgggt 3960
cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacgggt tatccacaga 4020
atcaggggat aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg 4080
taaaaaggcc gcgttgctgg cgtttttcca taggctccgc cccctgacg agcatcacia 4140
aaatcgacgc tcaagtcaaa ggtggcgaaa cccgacagga ctataaagat accaggcgtt 4200
tccccttgga agctccctcg tgcgtctctc ctgcccgtta ccggtaccc 4260
gtccgccttt ctcccttcgg gaagcgtggc gctttctcaa tgctcacgct gtaggtatct 4320
cagttcggtg taggtcgttc gctccaagct gggctgtgtg cacgaacccc ccgttcagcc 4380
cgaccgctgc gccttatccg gtaactatcg tcttgagtc aaccgggtta gacagactt 4440
atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg taggcgggtg 4500
tacaggttc ttgaagtggg ggccctaacta cggctacact agaaggacag tatttggtat 4560
ctgcgtcttg ctgaagccag ttaccttcgg ggtagctctt gatccggcaa 4620
acaaaccacc gctggtagcg gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa 4680
aaaaggatct caagaagatc ctttgatctt ttctacgggg tctgacgctc agtggaaacga 4740
aaactcacgt taagggtatt tggctcatgag attatcaaaa aggatcttca cctagatcct 4800
tttaaatcaa aaatgaagtt ttaaataaat cttaaagtata tatgagtaaa cttggtctga 4860
cagttaccaaa tgcttaataa gtgaggcacc tatctcagcg atctgtctat ttctgtctac 4920
catagttgcc tegtgtagat aactacgata cgggaggggt taccatctgg 4980
ccccagtgct gcaatgatac cgcgagaccc acgctcaccg gctccagatt tatcagcaat 5040
aaaccagcca gccggaaggg ccgagcgagc aagtggctct gcaactttat ccgctccat 5100
ccagttctatt aattgttgcc ggaagctag agtaaagtag tgcgagttta atagtgtg 5160
caacgttgtt gccattgcta caggcatcgt ggtgtcacgc tgcgctgttg gtatggcttc 5220
attcagctcc ggttcccaac gatcaaggcg agttacatga tccccatgt tgtgcaaaaa 5280
agcggttagc tccttcggtc ctccgatcgt tgtcagaagt aagttggccg cagtgttatc 5340
actcatgggt atggcagcac tgcataattc tcttactgtc atgccatccg taagatgctt 5400
ttctgtgact ggtgagtact caaccaagtc attctgagaa tagtgtatgc ggcgaccgag 5460
ttgctcttgc ccggcgtcaa tacgggataa caccgcgcca catagcagaa ctttaaaagt 5520
gctcatcatt ggaaaaagct cttcggggcg aaaactctca aggatcttac cgctgttgag 5580
atccagttcg atgtaaccca ctctgacacc caactgatct tcagatctt ttactttcac 5640
cagcgtttct gggtagcaa aaacaggaag gcaaaaagg gaataaggcg 5700
gacacggaaa tgttgaatac tcatactctt cctttttcaa tattattgaa gcatttatca 5760
gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata aacaaatagg 5820
ggttccgcgc acatttcccc gaaaagtgc acctg 5855

```

<210> 128

<211> 303

<212> DNA

<213> Artificial Sequence

<220>

<223> Human FER-1 Promoter

<400> 128

```

tccatgacaa agcacttttt gagcccaagc ccagcctagc tcgagctaaa cgggcacaga 60
gacgccaccg ctgtcccaga ggcagtcggc tacgggtccc cgctcccag ctcccagaga 120
gcgcgcgagg gctccagcg gcccccctc cccacagca ggggcggggg cccgcgccc 180
ccggaaggag cgggctcggg gcgggcggcg ctgattggcc ggggcggggc tgacgccgac 240
gcggctataa gagaccacaa gcgacccgca gggccagacg ttcttcgccc agagtccgggt 300
acc 303

```

-103-

<210> 129
<211> 6521
<212> DNA
<213> Artificial Sequence

<220>
<223> pIRES-BSR Plasmid

<400> 129
tcaatattgg ccattagcca tattattcat tggttatata gcataaatca atattggcta 60
ttggccattg catacgttgt atctatatca taatatattg gctcatgtcc 120
aatatgaccg ccatgttggc attgattatt gactagttaa taattacggg caattacggg 180
gtcattagtt catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc 240
gcctggctga ccgcccaacg acccccgcgc attgacgtca ataatgacgt atgttcccat 300
agtaacgcca atagggaactt tccattgacg tcaatgggtg gagtatttac ggtaaaactgc 360
ccacttggca gtacatcaag tgtatcatat gccaaagtcc cccctatttg acgtcaatga 420
cggtaaatgg ccgccttggc attatgcccc gtacatgacc ttacgggact ttccctacttg 480
gcagtacatc tacgtattag tcatcgctat taccatgggt atgcggtttt ggcagtacac 540
caatgggcgt ggatagcggg ttgactcacg gggatttcca agtctccacc ccattgacgt 600
caatgggggt ttgttttggc accaaaatca acgggacttt ccaaaatgtc gtaacaactg 660
cgatcgcccg ccccgttgac gcaaatgggc ggtaggcgtg tacgggtggga ggtctatata 720
agcagagctc gtttagtgaa ccgtcagatc actagaagct ttattgcggt agtttatcac 780
agttaaattg ctaacgcagt cagtgccttc gacacaacag tctcgaactt aagctgcagt 840
gactctctta aggttagcct gcagaagtgg gtctgtgagg actgggcagg taagtatcaa 900
ggttacaaga caggtttaag gagaccaata gaaactgggc ttgtcgagac agagaagact 960
cttgcgtttt tgataggcac ctatttggct tactgacatc cacttttgct ttctctccac 1020
agggtgtccac tcccagttca attacagctc ttaaggctag agtacttaac acgactcaat 1080
ataggctagc ctcgagaatt cacgcgtcga gcatgcatct agggcggcca attccgcccc 1140
tctccctccc ccccccctaa cgttactggc cgaagccgct tgggaataagg ccggtgtgcg 1200
tttgtctata tgtgattttc caccatattg ccgtcttttg gcaatgtgag gccccggaaa 1260
cctggccctg tcttcttgac gagcattcct aggggtcctt cccctctcgc caaaggaatg 1320
caaggctctg tgaatgtcgt gaaggaaagc gttcctctgg aagcttcttg aagacaaaac 1380
acgtctgtag cgacccttgg caggcagcgg aacccccac ctggcgacag gtgctctgc 1440
ggccaaaagc cacgtgtata agatacacct gcaaaaggcg cacaacccca gtgccacgtt 1500
gtgagttgga tagttgtgga aagagtcaaa tgggtctcct caagcgtatt caacaagggg 1560
ctgaaggatg ccagaagggt accccattgt atgggactcg atctggggcc tgggtgcaca 1620
tgctttacat gtgttttagt gaggttaaaa aaacgtctag gcccccgaa ccacggggac 1680
gtgggttttc tttgaaaaac acgatgataa gcttgccaca acccaccatg aaaacattta 1740
acatttctca acaagatcta gaattagtag gaattagcag agagaagatt caacttcttt 1800
atgaggataa taaacatcat gtgggagcgg caattcgtac gaaaacagga gaaatcattt 1860
cggcagtaga tattgaagcg tatataggac gactaaactgt ttgtgcagaa gccattgcga 1920
ttggtagtgc agtttcgaat ggacaaaagg attttgacac gattgtagct gttagacacc 1980
cttattctga cgaagtagat agactatgca ccagattgtt gagtggttaag tccttgtggg 2040
agttgatttc agactatgca ctatttccac ttagtgtaag agaaatgaat ggcaagttag 2100
tcaaaactac gattgaagaa ctcattccac tcaaatatac cggaaattaa aagttttacc 2160
ataccaagct tggcgggcgg ccgcttccct ttagtgaggg ttaatgcttc gagcagacat 2220
gataagatac attgatgagt ttggacaaac cacaactaga atgcagtga aaaaatgctt 2280
tatttgtgaa atttgtgatg ctattgcttt atttgaacc attataagct gcaataaaca 2340
agttaacaac aacaattgca ttcattttat gtttcagggt cagggggaga tgtgggaggt 2400
tttttaaagc aagtaaaacc tctacaaatg tggtaaaatc cgataaggat cgatccgggc 2460
tggcgtaata gcgaagaggc ccgcaccgat agcgccctag cgcccttccc aacagttgcg cagcctgaat 2520
ggcgaatgga cgcgccctgt agcggcgcat taagcgcggc ggggtgtggg gttacgcgca 2580
gcgtgaccgc tacacttgcc agcgccctag cgccgctcc tttcgctttc ttcccttcc 2640
ttctcgccac gttcgccggc ttccccgctc aagctctaaa tcgggggctc cctttagggt 2700
tccgatttag agctttacgg cacctcgacc gcaaaaaact tgatttgggt gatggttcac 2760
gtagtggggc atcgccctga tagacggttt ttcgcccttt gacgttggag tccacgttct 2820
ttaatagtgg actcttgttc caaactggaa caaactcaa ccctatctcg gtcattctt 2880
ttgatttata cggatttcgg ccgatttggg cctattgggt aaaaaatgag ctgatttaac 2940
aaatatttaa cgcgaatttt aacaaaatat taacgtttac aatttcgcct gatgcggtat 3000
tttctcctta cgcactctgt cggattttca caccgcatac gcggatctgc gcagcaccat 3060
ggcctgaaat aacctctgaa agagggaact ggttgaggt ctctgaggc ggaagaacc 3120
agctgtggaa tgtgtgtcag ttaggggtgt gaaagtcccc aggtctccca gcaggcagaa 3180
gtatgcaaag catgcatctc aattagtcag caaccagggt tggaaagtcc ccaggctccc 3240
cagcaggcag aagtagtcaa agcatgcatc tcaattagtc agcaaccata gtcccgcctc 3300
taactccgcc catccgccc ctaactccgc ccagttccgc ccctctcgct 3360
gactaatatt ttttatttat gcagaggccg aggcgcgctc ggcctctgag ctattccaga 3420
agtagtgagg aggccttttt ggaggccctag gcttttgcaa aaagcttgat tcttctgaca 3480

-104-

caacagtcctc	gaacttaagg	ctagagccac	catgattgaa	caagatggat	tgcacgcagg	3540
ttctccggcc	gcttgggtgg	agaggctatt	cggtatgac	tgggcacaa	agacaatcgg	3600
ctgctctgat	gccgcggtgt	tccggctgtc	agcgagggg	cgcccggttc	tttttgtcaa	3660
gaccgacctg	tccggtgccc	tgaatgaact	gcaggacgag	gcagcgcggc	tatcgtggct	3720
ggccacgacg	ggcggttcctt	gocgagctgt	gctcgacgtt	gtcactgaag	cggaagggga	3780
ctggctgcta	ttgggcgaag	tgccggggca	ggatctcctg	tcactctacc	ttgctcctgc	3840
cgagaaagta	tccatcatgg	ctgatgcaat	gcggcggtctg	catacgcttg	atccggctac	3900
ctgccccattc	gaccaccaag	cgaaacatcg	catcgagcga	gcacgtactc	ggatggaagc	3960
cggtcttgtc	gatcaggatg	atctggacga	agagcatcag	gggctcgcg	cagccgaact	4020
gttcgccagg	ctcaaggcgc	gcatgcccga	cggcgaggat	ctcgtcgtga	cccatggcga	4080
tgcctgcttg	ccgaatatca	tgggtgaaaa	tggccgcttt	tctggattca	tcgactgtgg	4140
ccggctgggt	gtggcgacc	gctatcagga	catagcgttg	gctaccctg	atattgctga	4200
agagcttggc	ggcgaatggg	ctgaccgctt	cctcgtgctt	tacggtatcg	ccgctcccga	4260
ttcgcagcgc	atcgcttctt	atcgcttctt	tgacgagttc	ttctgagcgg	gactctgggg	4320
ttcgaaatga	ccgaccaagc	gacgccccac	ctgccatcac	gatggcgcca	ataaaatata	4380
tttatttttca	ttacatctgt	gtgttgggtt	tttgtgtgaa	tcgatagcga	taaggatccg	4440
cgtatgggtgc	actctcagta	caatctgctc	tgatgccgca	tagttaagcc	agccccgaca	4500
cccggccaaca	cccgctgacg	cgccctgacg	ggcttgtctg	ctcccggcat	ccgcttacag	4560
acaagctgtg	accgtctccg	ggagctgcat	gtgtcagagg	ttttaccgt	catcacgcaa	4620
acgcgcgaga	cgaaagggcc	tcgtgatagc	cctattttta	taggttaatg	tcattgataat	4680
aatgttttct	tagacgtcag	gtggcacttt	tcggggaaat	gtgcgcggaa	cccctatttg	4740
tttattttttc	taaatacatt	caaatatgta	tccgctcatg	agacaataac	cctgataaat	4800
gcttcaataa	tattgaaaaa	ggaagagtat	gagtatccaa	catttccgtg	tcgccccttat	4860
tccctttttt	cgggcatttt	gccttccctg	ttttgctcac	ccagaaaacg	tggtgaaagt	4920
aaaagatgct	gaagatcagt	tgggtgcacg	agtgggttac	atcgaactgg	atctcaacag	4980
cggttaagatc	cttgagagtt	ttcgccccga	agaacgtttt	ccaatgatga	gcacttttaa	5040
agttctgcta	tgtggcgcg	tattatcccg	tattgacg	gggcaagagc	aactcggtcg	5100
ccgcatacac	tattctcaga	atgaactgg	tgagtactca	ccagtcacag	aaaagcatct	5160
tacggatggc	atgacagtaa	gagaattatg	cagtgtgccc	ataaccatga	gtgataacac	5220
tgcggccaac	ttacttctga	caacgatcgg	aggaccgaag	gagctaaccg	cttttttgca	5280
caacatgggg	gatcatgtaa	ctcgccctga	tcgttgggaa	ccggagctga	atgaagccat	5340
accaaaccgac	gagcgtgaca	ccacgatgcc	tgtagcaatg	gcaacaacgt	tcgcgaaaact	5400
attaactggc	gaactactta	ctctagcttc	ccggcaacaa	ttaatagact	ggatggaggc	5460
ggataaagtt	gcaggaccac	ttctgcgctc	ggcccttccg	gctggctgg	ttattgctga	5520
taaatctgga	gccgggtgag	gtgggtctcg	cggtatcatt	gcagcactgg	ggccagatgg	5580
taagccctcc	cgtatcgtag	ttatctacac	gacggggagt	caggcaacta	tgatgaacg	5640
aaatagacag	atcgctgaga	taggtgcctc	actgattaag	cattggtaac	tgtcagacca	5700
agtttactca	tatatacttt	agattgattt	aaaacttcat	ttttaattta	aaaggatcta	5760
ggtgaagatc	ctttttgata	atctcatgac	caaaatccct	taacgtgagt	tttcggtcca	5820
ctgagcgta	gaccccgtag	aaaagatcaa	aggatcttct	tgagatcctt	ttttctg	5880
cgtaatctgc	tgcttgcaaa	caaaaaaac	accgctacca	gcggtgggtt	gtttgcccga	5940
tcaagagcta	ccaactcttt	ttccgaaggt	aactggcttc	agcagagcgc	agataccaaa	6000
tactgtcctt	ctagtgtagc	cgtagttagg	ccaccacttc	aagaactctg	tagcaccgcc	6060
tacatacctc	gctctgctaa	tcctgttacc	agtggctgct	gccagtgccg	ataagtctgt	6120
tcttaccggg	ttggactcaa	gacgatagtt	accggataag	gcgacgcgg	cggtctgaac	6180
gggggggttc	tgcacacagc	ccagcttgga	gcgaacgacc	tacaccgaac	tgagatacct	6240
acagcgtgag	ctatgagaaa	gcgccacgct	tcccgaagg	agaaaggcgg	acaggtatcc	6300
ggttaagcgg	agggctcgaa	caggagagcg	cacgagggag	cttcacgggg	gaaacgcctg	6360
gtatctttat	agtcctgtcg	ggtttcgcca	cctctgactt	gagcgtcgat	ttttgtgatg	6420
ctcgtcaggg	gggcggagcc	tatggaaaaa	cgccagcaac	gcggcctttt	tacgggttct	6480
ggccttttgc	tggccttttg	ctcacatggc	tcgacagatc	t		6521

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number
WO 02/097059 A3

(51) International Patent Classification⁷: C07H 21/04,
C12N 15/85, 15/87, 15/90, A01N 43/34, C12N 15/09

(21) International Application Number: PCT/US02/17452

(22) International Filing Date: 30 May 2002 (30.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/294,758 30 May 2001 (30.05.2001) US
60/366,891 21 March 2002 (21.03.2002) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 60/294,758 (CIP)
Filed on 30 May 2001 (30.05.2001)
US 60/366,891 (CIP)
Filed on 21 March 2002 (21.03.2002)

(71) Applicant (for all designated States except US): CHROMOS MOLECULAR SYSTEMS, INC. [CA/CA]; 8081
Lougheed Highway, Burnaby, B.C. V5A 1W9 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): PERKINS, Edward [US/CA]; 7610 Lawrence Drive, Burnaby, B.C. V5A 1T6 (CA). PEREZ, Carl [US/CA]; 1201-7680 Granville Avenue, Richmond, B.C. V6Y 4B9 (CA). LINDENBAUM, Michael [CA/CA]; 252 Finnigan Street, Coquitlam, B.C. V3K 5J7 (CA). GREENE, Amy [US/CA]; 7610 Lawrence Drive, Burnaby, B.C. V5A 1T6 (CA). LEUNG, Josephine [CA/CA]; 711 Ebert Avenue, Coquitlam, B.C. V3J 7P8 (CA). FLEMING, Elena [CA/CA]; 248 E 18th, North Vancouver, B.C. V7L 2X6 (CA). STEWART, Sandra [CA/CA]; 2618 Oxford Street, Vancouver, B.C. V5K 1N3 (CA). SHELLARD, Joan [CA/CA]; #215-1345 West 15th Avenue, Vancouver, B.C. V6H 3R3 (CA).

(74) Agents: SEIDMAN, Stephanie, L. et al.; Heller Ehrman White & McAuliffe LLP, 7th floor, 4350 La Jolla Village Drive, San Diego, CA 92122-1246 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,

MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)
— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Published:

— with international search report

(88) Date of publication of the international search report:
30 May 2003

[Continued on next page]

(54) Title: CHROMOSOME-BASED PLATFORMS

(57) Abstract: Artificial chromosomes, including *Aces*, that have been engineered to contain available sites for site-specific, recombination-directed integration of DNA of interest are provided. These artificial chromosomes permit tractable, efficient, rational engineering of the chromosome for a variety of applications.

WO 02/097059 A3

WO 02/097059 A3



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/17452

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07H 21/04; C12N 15/85,87,90; A01N 43/34; C12N 15/09

US CL : 536/23.1; 435/320.1,325,419,455,467; 514/44; 800/21

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1; 435/320.1,325,419,455,467; 514/44; 800/21

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
APS (EAST); STN (MEDLINE, BIOSIS, CAPLUS)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 01/07572 A2 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 01 February 2001 (01.02.2001), see entire document.	1,2,7,8,11,23-25,27,50 ----- 3-6,9,10,12,13,26,28- 1,2,7,8,11,23-25,27,50 -----
X --- Y	WO 00/11155 A1 (THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY) 02 March 2000 (02.03.2000), see entire document.	3-6,9,10,12,13,26,28- 1,2,7,8,11,23-25,27,50 -----
X --- Y	US 6,171,861 B1 (HARTLEY et al.) 09 January 2001 (09.01.2001), see entire document, especially sequence listing.	3-6,9,10,12,13,26,28- 1-3,8,10-12,23-27,33,36,39,40,43-45,50,54,55,85 ----- 4-7,9,13-22,28-30,32,34,35,37,38,41,42,46-49,51-53,56,67-71,84,86 -----

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "B" earlier application or patent published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search

01 November 2002 (01.11.2002)

Date of mailing of the international search report

27 NOV 2002

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

Daniel M Sullivan

Telephone No. 703-308-0196

INTERNATIONAL SEARCH REPORT

PCT/US02/17452

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 96/40724 A1 (LIFE TECHNOLOGIES, INC.) 19 December 1996 (19.12.96), see entire document, especially SEQ ID NO: 1-16.	1-3,8,10-12,23- 27,33,36,39,40,43- 45,50,54,55,85 ----- 1-3,8,10-12,23- 27,33,36,39,40,43- 45,50,54,55,85
Y	WO 94/00569 A1 (GENPHARM INTERNATIONAL, INC.) 06 January 1994 (06.01.94), see entire document.	72-78
Y	WO 94/23049 A2 (THE JOHNS HOPKINS UNIVERSITY) 13 October 1994 (13.10.94), see entire document.	72-78
X --- Y	US 5,721,118 A (SCHEFFLER) 24 February 1998 (24.02.98), see entire document, especially Example 3.	14,15,22-25,57 ----- 16-21,26,58-64,72- 78,84,85,98-105,123
X --- Y	US 5,948,653 A (PATI et al.) 07 September 1999 (07.09.99), see entire document, especially the first full paragraph in column 29.	14,15,22-25,57 ----- 16-21,26,58-64,72- 78,84,85,98-105,123
X --- Y	GB 2 331 752 A (MEDICAL RESEARCH COUNCIL) 02 June 1999 (02.06.99), see entire document, especially pages 8-10.	109,123 ----- 72-78,110-122
A	HALDIMANN et al. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. J Bacteriol. November 2001, Vol. 183 No. 21, pages 6384-6393, see entire document.	65,66
X,P --- Y,P	MORALLI et al. Insertion of a loxP site in a size-reduced human accessory chromosome. Cytogenet. Cell Genet. 2001, Vol. 94 No. 3-4, pages 113-120, see entire document.	91-93,106,108 ----- 94-97,107
X --- Y	LORBACH et al. Site-specific recombination in human cells catalyzed by phage lambda integrase mutants. J. Mol. Biol. March 2000, Vol. 296 No. 5, pages 1175-1181, see entire document.	1,2,8,50 ----- 3-7,9-13,23-32,51- 56,84,85
X --- Y	CALL et al. A cre-lox recombination system for the targeted integration of circular yeast artificial chromosomes into embryonic stem cells. Hum. Mol. Genet. July 2000, Vol. 9, No. 12, pages 1745-1751, see entire document.	14,15,22-26,57,59 ----- 16- 21,58,60,84,85,98- 105
Y,P	HADLACZKY, G. Satellite DNA-based artificial chromosomes for use in gene therapy. Curr. Opin. Mol. Ther. April 2001, Vol. 3 No. 2, pages 125-132, see entire document.	1-123

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/17452

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐
☐

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

PCT/US02/17452

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-64, 67-71, 79, 84-86, 91-108, and 123, drawn to eukaryotic recombinogenic chromosomes used for introducing heterologous nucleic acids into a chromosome and the resulting cells.
Group II, claim(s) 65-66 and 87-88, drawn to a lambda intR mutain.
Group III, claim(s) 72-78, drawn to the production of transgenic animals.
Group IV, claim(s) 80, drawn to the production of an artificial chromosome library.
Group V, claim(s) 81-83, drawn to a library of cells for genomic screening.
Group VI, claim(s) 89-90, drawn to a modified iron-induced promoter.
Group VII, claim(s) 109-122, drawn to a method for screening compounds and their effects on regulatory regions.

and it considers that the International Application does not comply with the requirements of unity of invention (Rules 13.1, 13.2 and 13.3) for the reasons indicated below:

The inventions listed as Groups I-VII do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The special technical feature of Group I which defines an advance over the art is a eukaryotic chromosome containing recombinogenic sites that can be used to introduce heterologous nucleic acids into chromosomes, and cells containing these recombinogenic chromosomes.

The special technical feature of Group II involves a lambda intR mutain. This feature defines an advance over Group I in that it involves a protein that is not required for the technical features as set forth above in Group I.

The special technical feature of Group III involves the production of a transgenic animal, which represents a second method for using the invention as set forth above in Group I.

The special technical feature of Group IV involves the production of artificial chromosome expression system libraries, which represents a third method for using the invention as set forth above in Group I.

The special technical feature of Group V involves a library of cells containing the artificial chromosome expression system libraries set forth above in Group IV, and represents a first product resulting from Group IV.

The special technical feature of Group VI involves a modified iron-inducible promoter. This feature defines an advance over Group I in that it involves a promoter that is not required for the technical features as set forth above in Group I.

The special technical feature of Group VII involves a method for screening compounds for their effect on regulatory regions, which represents a fourth method of using the invention as set forth above in Group I.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number
WO 2002/097059 A3

(51) International Patent Classification⁷: **C07H 21/04**,
C12N 15/85, 15/87, 15/90, A01N 43/34, C12N 15/09

(21) International Application Number:
PCT/US2002/017452

(22) International Filing Date: 30 May 2002 (30.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/294,758 30 May 2001 (30.05.2001) US
60/366,891 21 March 2002 (21.03.2002) US

(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier applications:

US 60/294,758 (CIP)
Filed on 30 May 2001 (30.05.2001)
US 60/366,891 (CIP)
Filed on 21 March 2002 (21.03.2002)

(71) Applicant (for all designated States except US): **CHROMOS MOLECULAR SYSTEMS, INC.** [CA/CA]; 8081
Lougheed Highway, Burnaby, B.C. V5A 1W9 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **PERKINS, Edward**
[US/US]; 4203 Robinson Street, Duluth, MN 55804 (US). **PEREZ, Carl** [US/CA]; 95 West 11th Avenue, Vancouver, British Columbia V5Y 1S6 (CA). **LINDENBAUM, Michael** [CA/CA]; 9941 Martin Court, Burnaby, British Columbia V3K 1J5 (CA). **GREENE, Amy** [US/US]; 4203 Robinson Street, Duluth, MN 55804 (US). **LEUNG, Josephine** [CA/CA]; 711 Ebert Avenue, Coquitlam, B.C. V3J 7P8 (CA). **FLEMING, Elena** [CA/CA]; 427 Montroyal Blvd., North Vancouver, British Columbia V7N 3E2 (CA). **STEWART, Sandra** [CA/CA]; 2618 Oxford Street, Vancouver, B.C. V5K 1N3 (CA). **SHELLARD, Joan** [CA/CA]; #215-1345 West 15th Avenue, Vancouver, B.C. V6H 3R3 (CA).

(74) Agents: **SEIDMAN, Stephanie, L.** et al.; Heller Ehrman White & McAuliffe LLP, 7th floor, 4350 La Jolla Village Drive, San Diego, CA 92122-1246 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,

LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Published:

— with international search report
— with amended claims

(88) Date of publication of the international search report:
30 May 2003

[Continued on next page]

(54) Title: CHROMOSOME-BASED PLATFORMS

(57) Abstract: Artificial chromosomes, including *Aces*, that have been engineered to contain available sites for site-specific, recombination-directed integration of DNA of interest are provided. These artificial chromosomes permit tractable, efficient, rational engineering of the chromosome for a variety of applications.

WO 2002/097059 A3



Date of publication of the amended claims: 31 December 2003

(15) Information about Correction:

Previous Correction:

see PCT Gazette No. 39/2003 of 25 September 2003, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

AMENDED CLAIMS

[received by the International Bureau on 27 January 2003 (27.01.2003);
Original claims 22, 29, 39, 40, 68, 104, 107, 108, 114 and 121 replaced by
Amended Claims 22, 29, 39, 40, 68, 104, 107, 108, 114 and 121.;
Remaining claims unchanged]

13. The chromosome of claim 6 that is an artificial chromosome expression system (*ACes*).
14. A platform artificial chromosome expression system (*ACes*) comprising one or a plurality of sites that participate in recombinase catalyzed recombination.
15. The *ACes* of claim 14 that contains one site.
16. The *ACes* of claim 14 that is predominantly heterochromatin.
17. The *ACes* of claim 14 that contains no more than about 30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% euchromatin.
18. The *ACes* of claim 14 that is a plant *ACes*.
19. The *ACes* of claim 14 that is an animal *ACes*.
20. The *ACes* of claim 14 that is selected from a fish, insect, reptile, amphibian, arachnid or a mammalian *ACes*.
21. The *ACes* of claim 14 that is a fish *ACes*.
22. The artificial chromosome expression system (*ACes*) of claim 14, wherein the recombinase and site(s) are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase system of phage Mu, the Cin recombinase system, the Pin recombinase system of *E. coli*, the R/RS system of the pSR1 plasmid, or any combination thereof.
23. A method of introducing heterologous nucleic acid into a chromosome, comprising:
contacting a chromosome of any of claims 1 or 14 with a nucleic acid molecule comprising both the heterologous nucleic acid and a recombination site, in the presence of a recombinase that promotes recombination between the sites in the chromosome and in the nucleic acid molecule.

24. The method of claim 23, wherein the recombinase is selected from the group consisting of Cre, Gin, Cin, Pin, FLP, a phage integrase and R from the pSR1 plasmid.

25. The method of claim 23, wherein the nucleic acid molecule
5 encodes a therapeutic protein, antisense nucleic acid, or comprises an artificial chromosome.

26. The method of claim 25, wherein the nucleic acid molecule comprises a yeast artificial chromosomes (YAC), a bacterial artificial chromosome (BAC) or an insect artificial chromosome (IAC).

10 27. A combination, comprising, the chromosome of claim 1 and a first vector comprising the cognate recombination site, wherein the cognate recombination site is a site that recombines with the site engineered into the chromosome.

15 28. The combination of claim 27, further comprising nucleic acid encoding a recombinase, wherein the nucleic acid is on a second vector or on the first vector, or on the ACes under an inducible promoter.

29. The combination of claim 28, wherein the recombinase and sites are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase
20 system of phage Mu, the Pin recombinase system of *E. coli*, the R/RS system of the pSR1 plasmid, or any combination thereof.

30. The combination of claim 28, wherein a vector is the plasmid pCXLamIntR.

25 31. The combination of claim 27, wherein a vector is the plasmid pDsRedN1-attB.

32. A kit, comprising the combination of claim 27 and optionally instructions for introducing heterologous nucleic acid into the chromosome.

33. A method for introducing heterologous nucleic acid into a platform artificial chromosome, comprising:

- (a) mixing an artificial chromosome comprising at least a first recombination site and a vector comprising at least a second
5 recombination site and the heterologous nucleic acid;
(b) incubating the resulting mixture in the presence of at least one recombination protein under conditions whereby recombination between the first and second recombination sites is effected, thereby introducing the heterologous nucleic acid into the artificial chromosome.

10 34. The method of claim 33, wherein the artificial chromosome is an ACes.

35. The method of claim 33, wherein said mixing step (a) is conducted in cells ex vivo.

15 36. The method of claim 33, wherein said mixing step (a) is conducted extracellularly in an in vitro reaction mixture.

37. The method of claim 33, wherein the at least one recombination protein is encoded by a bacteriophage selected from the group consisting of bacteriophage lambda, phi 80, P22, P2, 186, P4 and P1.

20 38. The method of claim 37, wherein the at least one recombination protein is encoded by bacteriophage lambda, or mutants thereof.

39. The method of claim 33, wherein at least one recombination protein is selected from the group consisting of Int, IHF, Xis, Cre, $\gamma\delta$, Tn3
25 resolvase, Hin, Gin, Cin and Flp.

40. The method of claim 33, wherein the recombination sites are selected from the group consisting of att and lox P sites.

66. The lambda-intR mutein of claim 65, wherein the lambda-intR mutein comprises SEQ ID NO:37.

67. The method of claim 46 wherein the promoterless marker is transcriptionally downstream of the heterologous nucleic acid, wherein
5 the heterologous nucleic acid encodes a heterologous protein, and wherein the expression level of the selectable marker is transcriptionally linked to the expression level of the heterologous protein.

68. The method of claim 67, wherein the selectable marker and the heterologous nucleic acid are transcriptionally linked by the presence
10 of an IRES between them.

69. The method of claim 68, wherein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic or fluorescent.

70. The method of claim 69, wherein the selectable marker is selected from the group consisting of green fluorescent protein (GFP), red fluorescent protein (RFP), blue fluorescent protein (BFP), and *E. coli* histidinol dehydrogenase.

71. The method of claim 67 further comprising expressing the
20 heterologous protein and isolating the heterologous protein.

72. A method for producing a transgenic animal, comprising introducing a platform-ACes into an embryonic cell.

73. The method of claim 72, wherein the embryonic cell is a stem cell.

74. The method of claim 72, wherein the embryonic cell is in an
25 embryo.

75. The method of claim 72, wherein the platform-ACes comprises heterologous nucleic acid that encodes a therapeutic product.

a sequence of nucleotides that targets the vector to an amplifiable region of a chromosome.

92. The vector of claim 91, wherein the amplifiable region comprises heterochromatic nucleic acid.

5 93. The vector of claim 91, wherein the amplifiable region comprises rDNA.

94. The vector of claim 93, wherein the rDNA comprises an intergenic spacer.

95. The vector of claim 91, further comprising nucleic acid
10 encoding a selectable marker that is not operably associated with any promoter.

96. The vector of claim 91, wherein the chromosome is a mammalian chromosome.

97. The vector of claim 91, wherein the chromosome is a plant
15 chromosome.

98. A cell of claim 57 that is a plant cell, wherein the ACes platform is a MAC.

99. The plant cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

20 100. The plant cell of claim 99, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

101. A cell of claim 57 that is an animal cell, wherein the ACes platform is a plant artificial chromosome (PAC).

25 102. The cell of claim 101 that is a mammalian cell.

103. The cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

104. The cell of claim 102, wherein the PAC comprises transcriptional regulatory sequence of nucleotides derived from animals.

105. The cell of claim 104, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

106. A method, comprising:
5 introducing a vector of claim 91 into a cell;
 growing the cells; and
 selecting a cell comprising an artificial chromosome that comprises one or more repeat regions.

107. The method of claim 106, wherein a sufficient portion of the
10 vector integrates into a chromosome in the cell to result in amplification of chromosomal DNA.

108. The method of claim 106, wherein the artificial chromosome is an *ACes*.

109. A method for screening, comprising:
15 contacting a cell comprising a reporter *ACes* with test compounds or known compounds, wherein:

 the reporter *ACes* comprises one or a plurality of reporter constructs;

20 a reporter construct comprises a reporter gene in operative linkage with a regulatory region responsive to test or known compounds; and
 detecting any increase or decrease in signal output from the reporter, wherein a change in the signal is indicative of activity of the test or known compound on the regulatory region.

110. The method of claim 109, wherein the reporter is operatively
25 linked to a promoter that controls expression of a gene in a signal transduction pathway, whereby activation or reduction in the signal indicates that the pathway is activated or down-regulated by the test compound.

111. The method of claim 109, wherein the reporter in the construct encodes drug resistance or encodes a fluorescent protein.

112. The method of claim 111, wherein the fluorescent protein is selected from the group consisting of red, green and blue fluorescent
5 proteins.

113. The method of claim 109, wherein the ACes comprises a plurality of reporter-linked constructs, each with a different reporter, whereby the pathway(s) affected by the test compounds can be elucidated.

10 114. The method of claim 109, wherein a reporter is operatively linked to a promoter that is transcriptionally regulated in response to DNA damage, and the test compounds are genotoxicants.

115. The method of claim 114, wherein the DNA damage is induced by apoptosis, necrosis or cell-cycle perturbations.

15 116. The method of claim 114, wherein unknown compounds are screened to assess whether they are genotoxicants.

117. The method of claim 114, wherein the promoter is a cytochrome P450-profiled promoter.

20 118. The method of claim 114, wherein the cell is in a transgenic animal and toxicity is assessed in the animal.

119. The method of claim 109, wherein:

the cell is a patient cell sample; the patient has a disease;

the regulatory region is one targeted by a drug or drug regimen;

and

25 the method assesses the effectiveness of a treatment for the disease for the particular patient.

120. The method of claim 119, wherein the cell is a tumor cell.

121. The method of claim 109, wherein the cell is a stem cell or a progenitor cell, whereby expression of the reporter is operatively linked to

a regulatory region expressed in the cells to thereby identify stem cells or progenitor cells.

122. The method of claim 109, wherein the cell is in an animal;
and the method comprises whole-body imaging to monitor expression of
5 the reporter in the animal.

123. A reporter *ACes* comprises one or a plurality of reporter constructs, wherein the reporter construct comprises a reporter gene in operative linkage with a regulatory region responsive to test or known compounds.

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number
WO 02/097059 A3(51) International Patent Classification⁷: C07H 21/04,
C12N 15/85, 15/87, 15/90, A01N 43/34, C12N 15/09

(21) International Application Number: PCT/US02/17452

(22) International Filing Date: 30 May 2002 (30.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/294,758 30 May 2001 (30.05.2001) US
60/366,891 21 March 2002 (21.03.2002) US(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier applications:US 60/294,758 (CIP)
Filed on 30 May 2001 (30.05.2001)
US 60/366,891 (CIP)
Filed on 21 March 2002 (21.03.2002)(71) Applicant (for all designated States except US): CHROMOSOMOLECULAR SYSTEMS, INC. [CA/CA]; 8081
Lougheed Highway, Burnaby, B.C. V5A 1W9 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): PERKINS, Edward
[US/CA]; 7610 Lawrence Drive, Burnaby, B.C. V5A 1T6
(CA). PEREZ, Carl [US/CA]; 1201-7680 Granville Avenue,
Richmond, B.C. V6Y 4B9 (CA). LINDENBAUM,
Michael [CA/CA]; 252 Finnigan Street, Coquitlam, B.C.
V3K 5J7 (CA). GREENE, Amy [US/CA]; 7610 Lawrence
Drive, Burnaby, B.C. V5A 1T6 (CA). LEUNG, Josephine
[CA/CA]; 711 Ebert Avenue, Coquitlam, B.C. V3J 7P8
(CA). FLEMING, Elena [CA/CA]; 248 E 18th, North
Vancouver, B.C. V7L 2X6 (CA). STEWART, Sandra
[CA/CA]; 2618 Oxford Street, Vancouver, B.C. V5K 1N3
(CA). SHELLARD, Joan [CA/CA]; #215-1345 West
15th Avenue, Vancouver, B.C. V6H 3R3 (CA).(74) Agents: SEIDMAN, Stephanie, L. et al.; Heller Ehrman
White & McAuliffe LLP, 7th floor, 4350 La Jolla Village
Drive, San Diego, CA 92122-1246 (US).(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZM, ZW.(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,
NE, SN, TD, TG).

Declarations under Rule 4.17:

— as to the applicant's entitlement to claim the priority of the
earlier application (Rule 4.17(iii)) for the following design-
ations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY,
BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC,
EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN,
IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV,
MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH,
PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR,
TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent
(GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB,
GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ,
CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG)— as to the applicant's entitlement to claim the priority of the
earlier application (Rule 4.17(iii)) for the following design-
ations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY,
BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC,
EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN,
IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV,
MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH,
PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR,
TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent
(GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB,

[Continued on next page]

(54) Title: CHROMOSOME-BASED PLATFORMS

(57) Abstract: Artificial chromosomes, including *Aces*, that have been engineered to contain available sites for site-specific, re-combination-directed integration of DNA of interest are provided. These artificial chromosomes permit tractable, efficient, rational engineering of the chromosome for a variety of applications.

WO 02/097059 A3



GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Published:

— with international search report

(88) Date of publication of the international search report:
30 May 2003

(48) Date of publication of this corrected version:

25 September 2003

(15) Information about Correction:

see PCT Gazette No. 39/2003 of 25 September 2003, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

-1-

CHROMOSOME-BASED PLATFORMS**RELATED APPLICATIONS**

Benefit of priority to U.S. provisional application Serial No. 60/294,758, filed May 30, 2001, to Perkins, *et al.*, entitled "CHROMOSOME-BASED PLATFORMS" and to U.S. provisional application Serial No. 60/366,891, filed March 21, 2002, to Perkins, *et al.*, entitled

5 "CHROMOSOME-BASED PLATFORMS" is claimed. Where permitted, the subject matter of which are herein incorporated by reference in their entirety.

This application is related to Provisional Application No. 60/294,687, filed May 30, 2001, by CARL PEREZ AND STEVEN

10 FABIJANSKI entitled *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING PLANT ARTIFICIAL CHROMOSOMES* and to U.S. Provisional Application No. 60/296,329, filed June 4, 2001, by CARL PEREZ AND STEVEN FABIJANSKI entitled

15 *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING PLANT ARTIFICIAL CHROMOSOMES*. This application also is related to U.S. Provisional Application No. 60/294,758, filed May 30, 2001, by EDWARD PERKINS *et al.*, entitled *CHROMOSOME-BASED PLATFORMS* and to U.S. Provisional Application No. 60/366,891, filed

20 March 21, 2002, by EDWARD PERKINS *et al.*, entitled *CHROMOSOME-BASED PLATFORMS*. This application is also related to U.S. application Serial Nos. (attorney dkt nos. 24601-419 and 419PC), filed on the same day herewith, entitled *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS OF PREPARING PLANT ARTIFICIAL CHROMOSOMES* to Perez *et al.*.

25 This application is related to U.S. application Serial No. 08/695,191, filed August 7, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND*

METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES, now U.S. Patent No. 6,025,155. This application is also related to U.S. application Serial No. 08/682,080, filed July 15, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF*
5 *AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*, now U.S. Patent No. 6,077,697. This application is also related U.S. application Serial No. 08/629,822, filed April 10, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING*
10 *ARTIFICIAL CHROMOSOMES* (now abandoned), and is also related to copending U.S. application Serial No. 09/096,648, filed June 12, 1998, by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES* and to U.S. application Serial No.
15 09/835,682, April 10, 1997 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES* (now abandoned). This application is also related to copending U.S. application Serial No. 09/724,726, filed November 28, 2000, U.S. application Serial
20 No. 09/724,872, filed November 28, 2000, U.S. application Serial No. 09/724,693, filed November 28, 2000, U.S. application Serial No. 09/799,462, filed March 5, 2001, U.S. application Serial No. 09/836,911, filed April 17, 2001, and U.S. application Serial No. 10/125,767, filed April 17, 2002, each of which is by GYULA
25 HADLACZKY and ALADAR SZALAY, and is entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*. This application is also related to International PCT application No. WO 97/40183. Where permitted the

subject matter of each of these provisional applications, international applications, and applications is incorporated by reference in its entirety.

FIELD OF INVENTION

Artificial chromosomes, including *ACes*, that have been engineered
5 to contain available sites for site-specific, recombination-directed
integration of DNA of interest are provided. These artificial chromosomes
permit tractable, efficient, rational engineering of the chromosome.

BACKGROUND

Artificial chromosomes

10 A variety of artificial chromosomes for use in plants and animals,
particularly higher plants and animals are available. In particular, U.S.
Patent Nos. 6,025,155 and 6,077,697 provide heterochromatic artificial
chromosomes designated therein as satellite artificial chromosomes
(SATACs) and now designated artificial chromosome expression systems
15 (*ACes*). These chromosomes are prepared by introducing heterologous
DNA into a selected plant or animal cell under conditions that result in
integration into a region of the chromosome that leads to an amplification
event resulting in production of a dicentric chromosome. Subsequent
treatment and growth of cells with dicentric chromosomes, including
20 further amplifications, ultimately results in the artificial chromosomes
provided therein. In order to introduce a desired heterologous gene (or a
plurality of heterologous genes) into the artificial chromosome, the
process is repeated introducing the desired heterologous genes and
nucleic acids in the initial targeting step. This process is time consuming
25 and tedious. Hence, more tractable and efficient methods for introducing
heterologous nucleic acid molecules into artificial chromosomes,
particularly *ACes*, are needed.

Therefore, it is an object herein to provide engineered artificial chromosomes that permit tractable, efficient and rational engineering of artificial chromosomes.

SUMMARY OF THE INVENTION

5 Provided herein are artificial chromosomes that permit tractable, efficient and rational engineering thereof. In particular, the artificial chromosomes provided herein contain one or a plurality of loci (sites) for site-specific, recombination-directed integration of DNA. Thus, provided
10 herein are platform artificial chromosome expression systems ("platform *ACes*") containing single or multiple site-specific, recombination sites. The artificial chromosomes and *ACes* artificial chromosomes include plant and animal chromosomes. Any recombinase system that effects site-specific recombination is contemplated for use herein.

15 In one embodiment, chromosomes, including platform *ACes*, are provided that contain one or more lambda *att* sites designed for recombination-directed integration in the presence of lambda integrase, and that are mutated so that they do not require additional factors. Methods for preparing such chromosomes, vectors for use in the methods, and uses of the resulting chromosomes are also provided.

20 Platform *ACes* containing the recombination site(s) and methods for introducing heterologous nucleic acid into such sites and vectors therefor, are provided.

 Also provided herein is a bacteriophage lambda (λ) integrase site-specific recombination system.

25 Methods using recombinase mediated recombination target gene expression vectors and/or genes for insertion thereof into platform chromosomes and the resulting chromosomes are provided.

 Combinations and kits containing the combinations of vectors encoding a recombinase and integrase and primers for introduction of the

site recognized thereby are also provided. The kits optionally include instructions for performing site-directed integration or preparation of *ACes* containing such sites.

Also provided herein are mammalian and plant cells comprising the
5 artificial chromosomes and *ACes* described herein. The cells can be nuclear donor cells, stem cells, such as a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell or an embryonic stem cell.

Also provided is a lambda-intR mutein comprising a glutamic acid to arginine change at position 174 of wild-type lambda-integrase3. Also
10 provided are transgenic animals and methods for producing a transgenic animal, comprising introducing a *ACes* into an embryonic cell, such as a stem cell or embryo. The *ACes* can comprise heterologous nucleic acid that encodes a therapeutic product. The transgenic animal can be a fish, insect, reptile, amphibians, arachnid or mammal. In certain embodiments,
15 the *ACes* is introduced by cell fusion, lipid-mediated transfection by a carrier system, microinjection, microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer.

The platform *ACes*, including plant and animal *ACes*, such as MACs, provided herein can be introduced into cells, such as, but not
20 limited to, animal cells, including mammalian cells, and into plant cells. Hence plant cells that contain platform MACs, animal cells that contain platform PACs and other combinations of cells and platform *ACes* are provided.

DESCRIPTION OF FIGURES

25 **FIGURE 1** provides a diagram depicting creation of an exemplary *ACes* artificial chromosome prepared using methods detailed in U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183. In this exemplified embodiment, the nucleic acid is targeted to an acrocentric chromosome in an animal or plant, and the

heterologous nucleic acid includes a sequence-specific recombination site and marker genes.

FIGURE 2 provides a map of pWEPuro9K, which is a targeting vector derived from the vector pWE15 (GenBank Accession # X65279; SEQ ID No. 31). Plasmid pWE15 was modified by replacing the *Sal*I (Klenow filled)/*Sma*I neomycin resistance encoding fragment with the *Pvu*II/*Bam*HI (Klenow filled) puromycin resistance-encoding fragment (isolated from plasmid pPUR, Clontech Laboratories, Inc., Palo Alto, CA; GenBank Accession no. U07648; SEQ ID No. 30) resulting in plasmid pWEPuro. Subsequently a 9 Kb *Not*I fragment from the plasmid pFK161 (see Example 1, see, also Csonka *et al.* (2000) *Journal of Cell Science* 113:3207-32161; and SEQ ID NO: 118), containing a portion of the mouse *rDNA* region, was cloned into the *Not*I site of pWEPuro resulting in plasmid pWEPuro9K.

FIGURE 3 depicts construction of an ACes platform chromosome with a single recombination site, such as loxP sites or an *attP* or *attB* site. This platform ACes chromosome is an exemplary artificial chromosome with a single recombination site.

FIGURE 4 provides a map of plasmid pSV40-193attPsensePur.

FIGURE 5 depicts a method for formation of a chromosome platform with multiple recombination integration sites, such as *attP* sites.

FIGURE 6 sets forth the sequences of the core region of *attP*, *attB*, *attL* and *attR* (SEQ ID Nos. 33-36).

FIGURE 7 depicts insertional recombination of a vector encoding a marker gene, DsRed and an *attB* site with an artificial chromosome containing an *attP* site.

FIGURE 8 provides a map of plasmid pCXLamIntR (SEQ ID NO: 112), which includes the Lambda integrase (E174R)-encoding nucleic acid.

FIGURE 9 diagrammatically summarizes the platform technology; marker 1 permits selection of the artificial chromosomes containing the integration site; marker 2, which is promoterless in the target gene expression vector, permits selection of recombinants. Upon
5 recombination with the platform marker 2 is expressed under the control of a promoter resident on the platform.

FIGURE 10 provides the vector map for the plasmid p18attBZEO-5'6XHS4eGFP (SEQ ID NO: 116).

FIGURE 11 provides the vector map for the plasmid p18attBZEO-
10 3'6XHS4eGFP (SEQ ID NO: 115).

FIGURE 12 provides the vector map for the plasmid p18attBZEO-(6XHS4)2eGFP (SEQ ID NO: 110).

FIGURES 13 AND 14 depict the integration of a PCR product by site-specific recombination as set forth in Example 8.

15 FIGURE 15 provides the vector map for the plasmid pPACrDNA as set forth in Example 9.A.

DETAILED DESCRIPTION OF THE INVENTION

A. DEFINITIONS

Unless defined otherwise, all technical and scientific terms used
20 herein have the same meaning as is commonly understood by one of skill in the art to which the invention(s) belong. All patents, patent applications, published applications and publications, Genbank sequences, websites and other published materials referred to throughout the entire disclosure herein, unless noted otherwise, are incorporated by reference
25 in their entirety. Where reference is made to a URL or other such identifier or address, it understood that such identifiers can change and particular information on the internet can come and go, but equivalent information can be found by searching the internet. Reference thereto evidences the availability and public dissemination of such information.

As used herein, nucleic acid refers to single-stranded and/or double-stranded polynucleotides, such as deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), as well as analogs or derivatives of either RNA or DNA. Also included in the term "nucleic acid" are analogs of nucleic acids such as peptide nucleic acid (PNA), phosphorothioate DNA, and other such analogs and derivatives. When referring to probes or primers, optionally labeled, with a detectable label, such as a fluorescent or radiolabel, single-stranded molecules are contemplated. Such molecules are typically of a length such that they are statistically unique and of low copy number (typically less than 5, preferably less than 3) for probing or priming a library. Generally a probe or primer contains at least 14, 16 or 30 contiguous nucleotides of sequence complementary to or identical to a gene of interest. Probes and primers can be 10, 20, 30, 50, 100 or more nucleotides long.

As used herein, DNA is meant to include all types and sizes of DNA molecules including cDNA, plasmids and DNA including modified nucleotides and nucleotide analogs.

As used herein, nucleotides include nucleoside mono-, di-, and triphosphates. Nucleotides also include modified-nucleotides, such as, but are not limited to, phosphorothioate nucleotides and deazapurine nucleotides and other nucleotide analogs.

As used herein, heterologous or foreign DNA and RNA are used interchangeably and refer to DNA or RNA that does not occur naturally as part of the genome in which it is present or which is found in a location or locations and/or in amounts in a genome or cell that differ from that in which it occurs in nature. Heterologous nucleic acid is generally not endogenous to the cell into which it is introduced, but has been obtained from another cell or prepared synthetically. Generally, although not necessarily, such nucleic acid encodes RNA and proteins that are not

normally produced by the cell in which it is expressed. Any DNA or RNA that one of skill in the art would recognize or consider as heterologous or foreign to the cell in which it is expressed is herein encompassed by heterologous DNA. Heterologous DNA and RNA may also encode RNA or proteins that mediate or alter expression of endogenous DNA by affecting transcription, translation, or other regulatable biochemical processes.

Examples of heterologous DNA include, but are not limited to, DNA that encodes a gene product or gene product(s) of interest, introduced for purposes of modification of the endogenous genes or for production of an encoded protein. For example, a heterologous or foreign gene may be isolated from a different species than that of the host genome, or alternatively, may be isolated from the host genome but operably linked to one or more regulatory regions which differ from those found in the unaltered, native gene. Other examples of heterologous DNA include, but are not limited to, DNA that encodes traceable marker proteins, such as a protein that confers traits including, but not limited to, herbicide, insect, or disease resistance; traits, including, but not limited to, oil quality or carbohydrate composition. Antibodies that are encoded by heterologous DNA may be secreted or expressed on the surface of the cell in which the heterologous DNA has been introduced.

As used herein, operative linkage or operative association, or grammatical variations thereof, of heterologous DNA to regulatory and effector sequences of nucleotides, such as promoters, enhancers, transcriptional and translational stop sites, and other signal sequences refers to the relationship between such DNA and such sequences of nucleotides. For example, operative linkage of heterologous DNA to a promoter refers to the physical relationship between the DNA and the promoter such that the transcription of such DNA is initiated from the

promoter by an RNA polymerase that specifically recognizes, binds to and transcribes the DNA.

In order to optimize expression and/or *in vitro* transcription, it may be necessary to remove, add or alter 5' untranslated portions of the clones to eliminate extra, potential inappropriate alternative translation initiation (*i.e.*, start) codons or other sequences that may interfere with or reduce expression, either at the level of transcription or translation. Alternatively, consensus ribosome binding sites (see, *e.g.*, Kozak (1991) *J. Biol. Chem.* 266:19867-19870) can be inserted immediately 5' of the start codon and may enhance expression.

As used herein, a sequence complementary to at least a portion of an RNA, with reference to antisense oligonucleotides, means a sequence having sufficient complementarity to be able to hybridize with the RNA, preferably under moderate or high stringency conditions, forming a stable duplex. The ability to hybridize depends on the degree of complementarity and the length of the antisense nucleic acid. The longer the hybridizing nucleic acid, the more base mismatches it can contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex.

As used herein, regulatory molecule refers to a polymer of deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) or a polypeptide that is capable of enhancing or inhibiting expression of a gene.

As used herein, recognition sequences are particular sequences of nucleotides that a protein, DNA, or RNA molecule, or combinations thereof, (such as, but not limited to, a restriction endonuclease, a modification methylase and a recombinase) recognizes and binds. For example, a recognition sequence for Cre recombinase (see, *e.g.*, SEQ ID NO:58) is a 34 base pair sequence containing two 13 base pair inverted

repeats (serving as the recombinase binding sites) flanking an 8 base pair core and designated loxP (see, *e.g.*, Sauer (1994) *Current Opinion in Biotechnology* 5:521-527). Other examples of recognition sequences, include, but are not limited to, attB and attP, attR and attL and others (see, *e.g.*, SEQ ID Nos. 8, 41-56 and 72), that are recognized by the recombinase enzyme Integrase (see, SEQ ID Nos. 37 and 38 for the nucleotide and encoded amino acid sequences of an exemplary lambda phage integrase).

The recombination site designated attB is an approximately 33 base pair sequence containing two 9 base pair core-type Int binding sites and a 7 base pair overlap region; attP (SEQ ID No. 72) is an approximately 240 base pair sequence containing core-type Int binding sites and arm-type Int binding sites as well as sites for auxiliary proteins IHF, FIS, and Xis (see, *e.g.*, Landy (1993) *Current Opinion in Biotechnology* 3:699-707; see, SEQ ID Nos. 8 and 72).

As used herein, a recombinase is an enzyme that catalyzes the exchange of DNA segments at specific recombination sites. An integrase herein refers to a recombinase that is a member of the lambda (λ) integrase family.

As used herein, recombination proteins include excisive proteins, integrative proteins, enzymes, co-factors and associated proteins that are involved in recombination reactions using one or more recombination sites (see, Landy (1993) *Current Opinion in Biotechnology* 3:699-707). The recombination proteins used herein can be delivered to a cell via an expression cassette on an appropriate vector, such as a plasmid, and the like. In other embodiments, the recombination proteins can be delivered to a cell in protein form in the same reaction mixture used to deliver the desired nucleic acid, such as a platform ACes, donor target vectors, and the like.

As used herein the expression "lox site" means a sequence of nucleotides at which the gene product of the cre gene, referred to herein as Cre, can catalyze a site-specific recombination event. A LoxP site is a 34 base pair nucleotide sequence from bacteriophage P1 (see, 5 *e.g.*, Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398-3402). The LoxP site contains two 13 base pair inverted repeats separated by an 8 base pair spacer region as follows: (SEQ ID NO. 57):

ATAACTTCGTATA ATGTATGC TATACGAAGTTAT

E. coli DH5Δlac and yeast strain BSY23 transformed with plasmid pBS44 10 carrying two loxP sites connected with a LEU2 gene are available from the American Type Culture Collection (ATCC) under accession numbers ATCC 53254 and ATCC 20773, respectively. The lox sites can be isolated from plasmid pBS44 with restriction enzymes *EcoRI* and *SaI*, or *XhoI* and *BamHI*. In addition, a preselected DNA segment can be inserted 15 into pBS44 at either the *SaI* or *BamHI* restriction enzyme sites. Other lox sites include, but are not limited to, LoxB, LoxL, LoxC2 and LoxR sites, which are nucleotide sequences isolated from *E. coli* (see, *e.g.*, Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398). Lox sites can also be produced by a variety of synthetic techniques (see, *e.g.*, Ito *et al.* (1982) 20 *Nuc. Acid Res.* 10:1755 and Ogilvie *et al.* (1981) *Science* 270:270).

As used herein, the expression "cre gene" means a sequence of nucleotides that encodes a gene product that effects site-specific recombination of DNA in eukaryotic cells at lox sites. One cre gene can be isolated from bacteriophage P1 (see, *e.g.*, Abremski *et al.* (1983) *Cell* 25 32:1301-1311). *E. coli* DH1 and yeast strain BSY90 transformed with plasmid pBS39 carrying a cre gene isolated from bacteriophage P1 and a GAL1 regulatory nucleotide sequence are available from the American Type Culture Collection (ATCC) under accession numbers ATCC 53255

and ATCC 20772, respectively. The cre gene can be isolated from plasmid pBS39 with restriction enzymes *XhoI* and *Sall*.

As used herein, site-specific recombination refers to site-specific recombination that is effected between two specific sites on a single
5 nucleic acid molecule or between two different molecules that requires the presence of an exogenous protein, such as an integrase or recombinase.

For example, Cre-lox site-specific recombination can include the following three events:

- 10 a. deletion of a pre-selected DNA segment flanked by lox sites;
- b. inversion of the nucleotide sequence of a pre-selected DNA segment flanked by lox sites; and
- c. reciprocal exchange of DNA segments proximate to
15 lox sites located on different DNA molecules.

This reciprocal exchange of DNA segments can result in an integration event if one or both of the DNA molecules are circular. DNA segment refers to a linear fragment of single- or double-stranded deoxyribonucleic acid (DNA), which can be derived from any source.
20 Since the lox site is an asymmetrical nucleotide sequence, two lox sites on the same DNA molecule can have the same or opposite orientations with respect to each other. Recombination between lox sites in the same orientation results in a deletion of the DNA segment located between the two lox sites and a connection between the resulting ends of the original
25 DNA molecule. The deleted DNA segment forms a circular molecule of DNA. The original DNA molecule and the resulting circular molecule each contain a single lox site. Recombination between lox sites in opposite orientations on the same DNA molecule result in an inversion of the nucleotide sequence of the DNA segment located between the two lox

sites. In addition, reciprocal exchange of DNA segments proximate to lox sites located on two different DNA molecules can occur. All of these recombination events are catalyzed by the gene product of the cre gene. Thus, the Cre-lox system can be used to specifically delete, invert, or
5 insert DNA. The precise event is controlled by the orientation of lox DNA sequences, in *cis* the lox sequences direct the *Cre* recombinase to either delete (lox sequences in direct orientation) or invert (lox sequences in inverted orientation) DNA flanked by the sequences, while *in trans* the lox sequences can direct a homologous recombination event resulting in the
10 insertion of a recombinant DNA.

As used herein, a chromosome is a nucleic acid molecule, and associated proteins, that is capable of replication and segregation within a cell upon cell division. Typically, a chromosome contains a centromeric region, replication origins, telomeric regions and a region of nucleic acid
15 between the centromeric and telomeric regions.

As used herein, a centromere is any nucleic acid sequence that confers an ability to segregate to daughter cells through cell division. A centromere may confer stable segregation of a nucleic acid sequence, including an artificial chromosome containing the centromere, through
20 mitotic or meiotic divisions, including through both mitotic and meiotic divisions. A particular centromere is not necessarily derived from the same species in which it is introduced, but has the ability to promote DNA segregation in cells of that species.

As used herein, euchromatin and heterochromatin have their
25 recognized meanings. Euchromatin refers to chromatin that stains diffusely and that typically contains genes, and heterochromatin refers to chromatin that remains unusually condensed and that has been thought to be transcriptionally inactive. Highly repetitive DNA sequences (satellite DNA) are usually located in regions of the heterochromatin surrounding

the centromere (pericentric or pericentromeric heterochromatin). Constitutive heterochromatin refers to heterochromatin that contains the highly repetitive DNA which is constitutively condensed and genetically inactive.

5 As used herein, an acrocentric chromosome refers to a chromosome with arms of unequal length.

 As used herein, endogenous chromosomes refer to genomic chromosomes as found in a cell prior to generation or introduction of an artificial chromosome.

10 As used herein, artificial chromosomes are nucleic acid molecules, typically DNA, that stably replicate and segregate alongside endogenous chromosomes in cells and have the capacity to accommodate and express heterologous genes contained therein. It has the capacity to act as a gene delivery vehicle by accommodating and expressing foreign genes
15 contained therein. A mammalian artificial chromosome (MAC) refers to chromosomes that have an active mammalian centromere(s). Plant artificial chromosomes, insect artificial chromosomes and avian artificial chromosomes refer to chromosomes that include centromeres that function in plant, insect and avian cells, respectively. A human artificial
20 chromosome (HAC) refers to chromosomes that include centromeres that function in human cells. For exemplary artificial chromosomes, see, *e.g.*, U.S. Patent Nos. 6,025,155; 6,077,697; 5,288,625; 5,712,134; 5,695,967; 5,869,294; 5,891,691 and 5,721,118 and published International PCT application Nos, WO 97/40183 and WO 98/08964.
25 Artificial chromosomes include those that are predominantly heterochromatic (formerly referred to as satellite artificial chromosomes (SATACs); see, *e.g.*, U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183), minichromosomes that contain a *de novo* centromere (see, U.S. Patent

Nos. 5,712,134, 5,891,691 and 5,288,625), artificial chromosomes predominantly made up of repeating nucleic acid units and that contain substantially equivalent amounts of euchromatic and heterochromatic DNA and *in vitro* assembled artificial chromosomes (see, copending U.S. provisional application Serial No. 60/294,687, filed on May 30, 2001).

As used herein, the term "satellite DNA-based artificial chromosome (SATAC)" is interchangeable with the term "artificial chromosome expression system (ACes)". These artificial chromosomes (ACes) include those that are substantially all neutral non-coding sequences (heterochromatin) except for foreign heterologous, typically gene-encoding nucleic acid, that is interspersed within the heterochromatin for the expression therein (see U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183), or that is in a single locus as provided herein. Also included are ACes that may include euchromatin and that result from the process described in U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183 and outlined herein. The delineating structural feature is the presence of repeating units, that are generally predominantly heterochromatin. The precise structure of the ACes will depend upon the structure of the chromosome in which the initial amplification event occurs; all share the common feature of including a defined pattern of repeating units. Generally ACes have more heterochromatin than euchromatin. Foreign nucleic acid molecules (heterologous genes) contained in these artificial chromosome expression systems can include any nucleic acid whose expression is of interest in a particular host cell. Such foreign nucleic acid molecules, include, but are not limited to, nucleic acid that encodes traceable marker proteins (reporter genes), such as fluorescent proteins, such as green, blue or red fluorescent proteins (GFP, BFP and RFP, respectively), other reporter

genes, such as β -galactosidase and proteins that confer drug resistance, such as a gene encoding hygromycin-resistance. Other examples of heterologous nucleic acid molecules include, but are not limited to, DNA that encodes therapeutically effective substances, such as anti-cancer
5 agents, enzymes and hormones, DNA that encodes other types of proteins, such as antibodies, and DNA that encodes RNA molecules (such as antisense or siRNA molecules) that are not translated into proteins.

As used herein, an artificial chromosome platform, also referred to
10 herein as a "platform ACes" or "ACes platform", refers to an artificial chromosome that has been engineered to include one or more sites for site-specific, recombination-directed integration. In particular, ACes that are so-engineered are provided. Any sites, including but not limited to any described herein, that are suitable for such integration are
15 contemplated. Plant and animal platform ACes are provided. Among the ACes contemplated herein are those that are predominantly heterochromatic (formerly referred to as satellite artificial chromosomes (SATACs); see, *e.g.*, U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183), artificial
20 chromosomes predominantly made up of repeating nucleic acid units and that contain substantially equivalent amounts of euchromatic and heterochromatic DNA resulting from an amplification event depicted in the referenced patent and herein. Included among the ACes for use in generating platforms, are artificial chromosomes that introduce and
25 express heterologous nucleic acids in plants (see, copending U.S. provisional application Serial No. 60/294,687, filed on May 30, 2001). These include artificial chromosomes that have a centromere derived from a plant, and, also, artificial chromosomes that have centromeres that may be derived from other organisms but that function in plants.

As used herein a "reporter ACes" refers to an ACes that comprises one or a plurality of reporter constructs, where the reporter construct comprises a reporter gene in operative linkage with a regulatory region responsive to test or known compounds.

5 As used herein, amplification, with reference to DNA, is a process in which segments of DNA are duplicated to yield two or multiple copies of substantially similar or identical or nearly identical DNA segments that are typically joined as substantially tandem or successive repeats or inverted repeats.

10 As used herein, amplification-based artificial chromosomes are artificial chromosomes derived from natural or endogenous chromosomes by virtue of an amplification event, such as one initiated by introduction of heterologous nucleic acid into rDNA in a chromosome. As a result of such an event, chromosomes and fragments thereof exhibiting segmented
15 or repeating patterns arise. Artificial chromosomes can be formed from these chromosomes and fragments. Hence, amplification-based artificial chromosomes refer to engineered chromosomes that exhibit an ordered segmentation that is not observed in naturally occurring chromosomes and that distinguishes them from naturally occurring chromosomes. The
20 segmentation, which can be visualized using a variety of chromosome analysis techniques known to those of skill in the art, correlates with the structure of these artificial chromosomes. In addition to containing one or more centromeres, the amplification-based artificial chromosomes, throughout the region or regions of segmentation are predominantly made
25 up of nucleic acid units also referred to as "amplicons", that is (are) repeated in the region and that have a similar gross structure. Repeats of an amplicon tend to be of similar size and share some common nucleic acid sequences. For example, each repeat of an amplicon may contain a replication site involved in amplification of chromosome segments and/or

some heterologous nucleic acid that was utilized in the initial production of the artificial chromosome. Typically, the repeating units are substantially similar in nucleic acid composition and may be nearly identical.

5 The amplification-based artificial chromosomes differ depending on the chromosomal region that has undergone amplification in the process of artificial chromosome formation. The structures of the resulting chromosomes can vary depending upon the initiating event and/or the conditions under which the heterologous nucleic acid is introduced,
10 including modification to the endogenous chromosomes. For example, in some of the artificial chromosomes provided herein, the region or regions of segmentation may be made up predominantly of heterochromatic DNA. In other artificial chromosomes provided herein, the region or regions of segmentation may be made up predominantly of euchromatic DNA or may
15 be made up of similar amounts of heterochromatic and euchromatic DNA.

As used herein an amplicon is a repeated nucleic acid unit. In some of the artificial chromosomes described herein, an amplicon may contain a set of inverted repeats of a megareplicon. A megareplicon represents a higher order replication unit. For example, with reference to
20 some of the predominantly heterochromatic artificial chromosomes, the megareplicon can contain a set of tandem DNA blocks (*e.g.*, ~7.5 Mb DNA blocks) each containing satellite DNA flanked by non-satellite DNA or may be made up of substantially rDNA. Contained within the megareplicon is a primary replication site, referred to as the
25 megareplicator, which may be involved in organizing and facilitating replication of the pericentric heterochromatin and possibly the centromeres. Within the megareplicon there may be smaller (*e.g.*, 50-300 kb) secondary replicons.

In artificial chromosomes, such as those provided U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183, the megareplicon is defined by two tandem blocks (~7.5 Mb DNA blocks in the chromosomes provided therein). Within each artificial chromosome or among a population thereof, each amplicon has the same gross structure but may contain sequence variations. Such variations will arise as a result of movement of mobile genetic elements, deletions or insertions or mutations that arise, particularly in culture. Such variation does not affect the use of the artificial chromosomes or their overall structure as described herein.

As used herein, amplifiable, when used in reference to a chromosome, particularly the method of generating artificial chromosomes provided herein, refers to a region of a chromosome that is prone to amplification. Amplification typically occurs during replication and other cellular events involving recombination (*e.g.*, DNA repair). Such regions include regions of the chromosome that contain tandem repeats, such as satellite DNA, rDNA, and other such sequences.

As used herein, a dicentric chromosome is a chromosome that contains two centromeres. A multicentric chromosome contains more than two centromeres.

As used herein, a formerly dicentric chromosome is a chromosome that is produced when a dicentric chromosome fragments and acquires new telomeres so that two chromosomes, each having one of the centromeres, are produced. Each of the fragments is a replicable chromosome. If one of the chromosomes undergoes amplification of primarily euchromatic DNA to produce a fully functional chromosome that is predominantly (at least more than 50%) euchromatin, it is a minichromosome. The remaining chromosome is a formerly dicentric chromosome. If one of the chromosomes undergoes amplification,

whereby heterochromatin (such as, for example, satellite DNA) is amplified and a euchromatic portion (such as, for example, an arm) remains, it is referred to as a sausage chromosome. A chromosome that is substantially all heterochromatin, except for portions of heterologous DNA, is called a predominantly heterochromatic artificial chromosome. Predominantly heterochromatic artificial chromosomes can be produced from other partially heterochromatic artificial chromosomes by culturing the cell containing such chromosomes under conditions such as BrdU treatment that destabilize the chromosome and/or growth under selective conditions so that a predominantly heterochromatic artificial chromosome is produced. For purposes herein, it is understood that the artificial chromosomes may not necessarily be produced in multiple steps, but may appear after the initial introduction of the heterologous DNA. Typically, artificial chromosomes appear after about 5 to about 60, or about 5 to about 55, or about 10 to about 55 or about 25 to about 55 or about 35 to about 55 cell doublings after initiation of artificial chromosome generation, or they may appear after several cycles of growth under selective conditions and BrdU treatment.

As used herein, an artificial chromosome that is predominantly heterochromatic (*i.e.*, containing more heterochromatin than euchromatin, typically more than about 50%, more than about 70%, or more than about 90% heterochromatin) may be produced by introducing nucleic acid molecules into cells, such as, for example, animal or plant cells, and selecting cells that contain a predominantly heterochromatic artificial chromosome. Any nucleic acid may be introduced into cells in such methods of producing the artificial chromosomes. For example, the nucleic acid may contain a selectable marker and/or optionally a sequence that targets nucleic acid to the pericentric, heterochromatic region of a chromosome, such as in the short arm of acrocentric chromosomes and

nucleolar organizing regions. Targeting sequences include, but are not limited to, lambda phage DNA and rDNA for production of predominantly heterochromatic artificial chromosomes in eukaryotic cells.

After introducing the nucleic acid into cells, a cell containing a
5 predominantly heterochromatic artificial chromosome is selected. Such
cells may be identified using a variety of procedures. For example,
repeating units of heterochromatic DNA of these chromosomes may be
discerned by G-banding and/or fluorescence in situ hybridization (FISH)
techniques. Prior to such analyses, the cells to be analyzed may be
10 enriched with artificial chromosome-containing cells by sorting the cells
on the basis of the presence of a selectable marker, such as a reporter
protein, or by growing (culturing) the cells under selective conditions. It
is also possible, after introduction of nucleic acids into cells, to select
cells that have a multicentric, typically dicentric, chromosome, a formerly
15 multicentric (typically dicentric) chromosome and/or various
heterochromatic structures, such as a megachromosome and a sausage
chromosome, that contain a centromere and are predominantly
heterochromatic and to treat them such that desired artificial
chromosomes are produced. Cells containing a new chromosome are
20 selected. Conditions for generation of a desired structure include, but are
not limited to, further growth under selective conditions, introduction of
additional nucleic acid molecules and/or growth under selective conditions
and treatment with destabilizing agents, and other such methods (see
International PCT application No. WO 97/40183 and U.S. Patent Nos.
25 6,025,155 and 6,077,697).

As used herein, a "selectable marker" is a nucleic acid segment,
generally DNA, that allows one to select for or against a molecule or a
cell that contains it, often under particular conditions. These markers can
encode an activity, such as, but not limited to, production of RNA,

peptide, or protein, or can provide a binding site for RNA, peptides, proteins, inorganic and organic compounds and compositions. Examples of selectable markers include but are not limited to: (1) nucleic acid segments that encode products that provide resistance against otherwise
5 toxic compounds (e.g., antibiotics); (2) nucleic acid segments that encode products that are otherwise lacking in the recipient cell (e.g., tRNA genes, auxotrophic markers); (3) nucleic acid segments that encode products that suppress the activity of a gene product; (4) nucleic acid segments that encode products that can be identified, such as phenotypic markers,
10 including β -galactosidase, red, blue and/or green fluorescent proteins (FPs), and cell surface proteins; (5) nucleic acid segments that bind products that are otherwise detrimental to cell survival and/or function; (6) nucleic acid segments that otherwise inhibit the activity of any of the nucleic acid segments described in Nos. 1-5 above (e.g., antisense
15 oligonucleotides or siRNA molecules for use in RNA interference); (7) nucleic acid segments that bind products that modify a substrate (e.g. restriction endonucleases); (8) nucleic acid segments that can be used to isolate a desired molecule (e.g. specific protein binding sites); (9) nucleic acid segments that encode a specific nucleotide sequence that can be
20 otherwise non-functional, such as for PCR amplification of subpopulations of molecules; and/or (10) nucleic acid segments, which when absent, directly or indirectly confer sensitivity to particular compounds. Thus, for example, selectable markers include nucleic acids encoding fluorescent proteins, such as green fluorescent proteins, β -galactosidase and other
25 readily detectable proteins, such as chromogenic proteins or proteins capable of being bound by an antibody and FACs sorted. Selectable markers such as these, which are not required for cell survival and/or proliferation in the presence of a selection agent, are also referred to herein as reporter molecules. Other selectable markers, e.g., the

neomycin phosphotransferase gene, provide for isolation and identification of cells containing them by conferring properties on the cells that make them resistant to an agent, *e.g.*, a drug such as an antibiotic, that inhibits proliferation of cells that do not contain the marker.

- 5 As another example, interference of gene expression by double stranded RNA has been shown in *Caenorhabditis elegans*, plants, *Drosophila*, protozoans and mammals. This method is known as RNA interference (RNAi) and utilizes short, double-stranded RNA molecules (siRNAs). The siRNAs are generally composed of a 19-22bp double-
- 10 stranded RNA stem, a loop region and a 1-4 bp overhang on the 3' end. The reduction of gene expression has been accomplished by direct introduction of the siRNAs into the cell (Harborth J et al., 2001, J Cell Sci 114(pt 24):4557-65) as well as the introduction of DNA encoding and expressing the siRNA molecule. The encoded siRNA molecules are under
- 15 the regulation of an RNA polymerase III promoter (see, *e.g.*, Yu et al., 2002, Proc Natl Acad Sci USA 99(9):6047-52; Brummelkamp et al., 2002, Science 296(5567):550-3; Miyagishi et al., 2002, Nat Biotechnol 20(5):497-500; and the like). In certain embodiments, RNAi in mammalian cells may have advantages over other therapeutic methods.
- 20 For example, producing siRNA molecules that block viral genetic activities in infected cells may reduce the effects of the virus. Platform ACes provided herein encoding siRNA molecule(s) are an additional utilization of the platform ACes technology. The platform ACes could be engineered to encode one or more siRNA molecules to create gene "knockdowns". In
- 25 one embodiment, a platform ACes can be engineered to encode both the siRNA molecule and a replacement gene. For example, a mouse model or cell culture system could be generated using a platform ACes that has a knockdown of the endogenous mouse gene, by siRNA, and the human gene homolog expressing in place of the mouse gene. The placement of

siRNA encoding sequences under the regulation of a regulatable or inducible promoter would allow one to temporally and/or spatially control the knockdown effect of the corresponding gene.

As used herein, a reporter gene includes any gene that expresses a detectable gene product, which may be RNA or protein. Generally reporter genes are readily detectable. Examples of reporter genes include, but are not limited to nucleic acid encoding a fluorescent protein, CAT (chloramphenicol acetyl transferase) (Alton *et al.* (1979) *Nature* 282: 864-869) luciferase, and other enzyme detection systems, such as beta-galactosidase; firefly luciferase (deWet *et al.* (1987) *Mol. Cell. Biol.* 7:725-737); bacterial luciferase (Engebrecht and Silverman (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81:4154-4158; Baldwin *et al.* (1984) *Biochemistry* 23:3663-3667); and alkaline phosphatase (Toh *et al.* (1989) *Eur. J. Biochem.* 182:231-238, Hall *et al.* (1983) *J. Mol. Appl. Gen.* 2:101).

As used herein, growth under selective conditions means growth of a cell under conditions that require expression of a selectable marker for survival.

As used herein, an agent that destabilizes a chromosome is any agent known by those skilled in the art to enhance amplification events, and/or mutations. Such agents, which include BrdU, are well known to those skilled in the art.

In order to generate an artificial chromosome containing a particular heterologous nucleic acid of interest, it is possible to include the nucleic acid in the nucleic acid that is being introduced into cells to initiate production of the artificial chromosome. Thus, for example, a nucleic acid can be introduced into a cell along with nucleic acid encoding a selectable marker and/or a nucleic acid that targets to a heterochromatic region of a chromosome. For introducing a heterologous nucleic acid into

the cell, it can be included in a fragment that includes a selectable marker or as part of a separate nucleic acid fragment and introduced into the cell with a selectable marker during the process of generating the artificial chromosomes. Alternatively, heterologous nucleic acid can be introduced
5 into an artificial chromosome at a later time after the initial generation of the artificial chromosome.

As used herein, the minichromosome refers to a chromosome derived from a multicentric, typically dicentric, chromosome that contains more euchromatic than heterochromatic DNA. For purposes herein, the
10 minichromosome contains a *de novo* centromere (e.g., a neocentromere). In some embodiments, for example, the minichromosome contains a centromere that replicates in animals, e.g., a mammalian centromere or in plants, e.g., a plant centromere.

As used herein, *in vitro* assembled artificial chromosomes or
15 synthetic chromosomes can be either more euchromatic than heterochromatic or more heterochromatic than euchromatic and are produced by joining essential components of a chromosome *in vitro*. These components include at least a centromere, a megareplicator, a telomere and optionally secondary origins of replication.

20 As used herein, *in vitro* assembled plant or animal artificial chromosomes are produced by joining essential components (at least the centromere, telomere(s), megareplicator and optional secondary origins of replication) that function in plants or animals. In particular embodiments, the megareplicator contains sequences of rDNA, particularly plant or
25 animal rDNA.

As used herein, a plant is a eukaryotic organism that contains, in addition to a nucleus and mitochondria, chloroplasts capable of carrying out photosynthesis. A plant can be unicellular or multicellular and can contain multiple tissues and/or organs. Plants can reproduce sexually or

asexually and can be perennial or annual in growth. Plants can also be terrestrial or aquatic. The term "plant" includes a whole plant, plant cell, plant protoplast, plant calli, plant seed, plant organ, plant tissue, and other parts of a whole plant.

5 As used herein, stable maintenance of chromosomes occurs when at least about 85%, preferably 90%, more preferably 95%, of the cells retain the chromosome. Stability is measured in the presence of a selective agent. Preferably these chromosomes are also maintained in the absence of a selective agent. Stable chromosomes also retain their
10 structure during cell culturing, suffering no unintended intrachromosomal or interchromosomal rearrangements.

 As used herein, *de novo* with reference to a centromere, refers to generation of an excess centromere in a chromosome as a result of incorporation of a heterologous nucleic acid fragment using the methods
15 herein.

 As used herein, BrdU refers to 5-bromodeoxyuridine, which during replication is inserted in place of thymidine. BrdU is used as a mutagen; it also inhibits condensation of metaphase chromosomes during cell division.

20 As used herein, ribosomal RNA (rRNA) is the specialized RNA that forms part of the structure of a ribosome and participates in the synthesis of proteins. Ribosomal RNA is produced by transcription of genes which, in eukaryotic cells, are present in multiple copies. In human cells, the approximately 250 copies of rRNA genes (i.e., genes which encode rRNA)
25 per haploid genome are spread out in clusters on at least five different chromosomes (chromosomes 13, 14, 15, 21 and 22). In mouse cells, the presence of ribosomal DNA (rDNA, which is DNA containing sequences that encode rRNA) has been verified on at least 11 pairs out of 20 mouse chromosomes (chromosomes 5, 6, 7, 9, 11, 12, 15, 16, 17, 18, and 19)

(see e.g., Rowe *et al.* (1996) *Mamm. Genome* 7:886-889 and Johnson *et al.* (1993) *Mamm. Genome* 4:49-52). In *Arabidopsis thaliana* the presence of rDNA has been verified on chromosomes 2 and 4 (18S, 5.8S, and 25S rDNA) and on chromosomes 3,4, and 5 (5S rDNA)(see The
5 Arabidopsis Genome Initiative (2000) *Nature* 408:796-815). In eukaryotic cells, the multiple copies of the highly conserved rRNA genes are located in a tandemly arranged series of rDNA units, which are generally about 40-45 kb in length and contain a transcribed region and a nontranscribed region known as spacer (i.e., intergenic spacer) DNA
10 which can vary in length and sequence. In the human and mouse, these tandem arrays of rDNA units are located adjacent to the pericentric satellite DNA sequences (heterochromatin). The regions of these chromosomes in which the rDNA is located are referred to as nucleolar organizing regions (NOR) which loop into the nucleolus, the site of
15 ribosome production within the cell nucleus.

As used herein, a megachromosome refers to a chromosome that, except for introduced heterologous DNA, is substantially composed of heterochromatin. Megachromosomes are made up of an array of repeated amplicons that contain two inverted megareplicons bordered by
20 introduced heterologous DNA (see, e.g., Figure 3 of U.S. Patent No. 6,077,697 for a schematic drawing of a megachromosome). For purposes herein, a megachromosome is about 50 to 400 Mb, generally about 250-400 Mb. Shorter variants are also referred to as truncated megachromosomes (about 90 to 120 or 150 Mb), dwarf
25 megachromosomes (~150-200 Mb), and a micro-megachromosome (~50-90 Mb, typically 50-60 Mb). For purposes herein, the term

megachromosome refers to the overall repeated structure based on an array of repeated chromosomal segments (amplicons) that contain two inverted megareplicons bordered by any inserted heterologous DNA. The size will be specified.

5 As used herein, gene therapy involves the transfer or insertion of nucleic acid molecules into certain cells, which are also referred to as target cells, to produce specific products that are involved in preventing, curing, correcting, controlling or modulating diseases, disorders and deleterious conditions. The nucleic acid is introduced into the selected
10 target cells in a manner such that the nucleic acid is expressed and a product encoded thereby is produced. Alternatively, the nucleic acid may in some manner mediate expression of DNA that encodes a therapeutic product. This product may be a therapeutic compound, which is produced in therapeutically effective amounts or at a therapeutically
15 useful time. It may also encode a product, such as a peptide or RNA, that in some manner mediates, directly or indirectly, expression of a therapeutic product. Expression of the nucleic acid by the target cells within an organism afflicted with a disease or disorder thereby provides for modulation of the disease or disorder. The nucleic acid encoding the
20 therapeutic product may be modified prior to introduction into the cells of the afflicted host in order to enhance or otherwise alter the product or expression thereof.

 For use in gene therapy, cells can be transfected *in vitro*, followed by introduction of the transfected cells into an organism. This is often
25 referred to as *ex vivo* gene therapy. Alternatively, the cells can be transfected directly *in vivo* within an organism.

 As used herein, therapeutic agents include, but are not limited to, growth factors, antibodies, cytokines, such as tumor necrosis factors and interleukins, and cytotoxic agents and other agents disclosed herein and

known to those of skill in the art. Such agents include, but are not limited to, tumor necrosis factor, α -interferon, β -interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin- 1 (IL-1), interleukin-2 (IL-2), interleukin-6 (IL-6), granulocyte macrophage colony stimulating factor (GMCSF), granulocyte colony stimulating factor (G-CSF), erythropoietin (EPO), pro-coagulants such as tissue factor and tissue factor variants, pro-apoptotic agents such FAS-ligand, fibroblast growth factors (FGF), nerve growth factor and other growth factors.

As used herein, a therapeutically effective product is a product that is encoded by heterologous DNA that, upon introduction of the DNA into a host, a product is expressed that effectively ameliorates or eliminates the symptoms, manifestations of an inherited or acquired disease or that cures the disease.

As used herein, transgenic plants and animals refer to plants and animals in which heterologous or foreign nucleic acid is expressed or in which the expression of a gene naturally present in the plant or animal has been altered by virtue of introduction of heterologous or foreign nucleic acid.

As used herein, IRES (internal ribosome entry site; see, *e.g.*, SEQ ID No. 27 and nucleotides 2736-3308 SEQ ID No. 28) refers to a region of a nucleic acid molecule, such as an mRNA molecule, that allows internal ribosome entry sufficient to initiate translation, which initiation can be detected in an assay for cap-independent translation (see, *e.g.*, U.S. Patent No. 6,171,821). The presence of an IRES within an mRNA molecule allows cap-independent translation of a linked protein-encoding sequence that otherwise would not be translated.

Internal ribosome entry site (IRES) elements were first identified in picornaviruses, which elements are considered the paradigm for cap-independent translation. The 5' UTRs of all picornaviruses are long and mediate translational initiation by directly recruiting and binding
5 ribosomes, thereby circumventing the initial cap-binding step. IRES elements are frequently found in viral mRNA, they are rare in non-viral mRNA. Among non-viral mRNA molecules that contain functional IRES elements in their respective 5' UTRs are those encoding immunoglobulin heavy chain binding protein (BiP) (Macejak *et al.* (1991) *Nature*
10 353:90-94); *Drosophila* Antennapedia (Oh *et al.* (1992) *Genes Dev.* 6:1643-1653); *D. Ultrabithorax* (Ye *et al.* (1997) *Mol. Cell Biol.* 17:1714-21); fibroblast growth factor 2 (Vagner *et al.* (1995) *Mol. Cell Biol.* 15:35-44); initiation factor eIF4G (Gan *et al.* (1998) *J. Biol. Chem.* 273:5006-5012); proto-oncogene c-myc (Nanbru *et al.* (1995) *J. Biol.*
15 *Chem.* 272:32061-32066; Stoneley (1998) *Oncogene* 16:423-428); IRES_H; from the 5'UTR of NRF1 gene (Oumard *et al.* (2000) *Mol. and Cell Biol.*, 20(8):2755-2759); and vascular endothelial growth factor (VEGF) (Stein *et al.* (1998) *Mol. Cell Biol.* 18:3112-9).

As used herein, a promoter, with respect to a region of DNA, refers
20 to a sequence of DNA that contains a sequence of bases that signals RNA polymerase to associate with the DNA and initiate transcription of RNA (such as pol II for mRNA) from a template strand of the DNA. A promoter thus generally regulates transcription of DNA into mRNA. A particular promoter provided herein is the Ferritin heavy chain promoter (excluding
25 the Iron Response Element, located in the 5'UTR), which was joined to the 37bp Fer-1 enhancer element. This promoter is set forth as SEQ ID NO:128. The endogenous Fer-1 enhancer element is located upstream of the Fer-1 promoter (e.g., a Fer-1 oligo was cloned proximal to the core promoter).

As used herein, isolated, substantially pure nucleic acid, such as, for example, DNA, refers to nucleic acid fragments purified according to standard techniques employed by those skilled in the art, such as that found in Sambrook *et al.* ((2001) Molecular Cloning: A Laboratory
5 Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 3rd edition).

As used herein, expression refers to the transcription and/or translation of nucleic acid. For example, expression can be the transcription of a gene that may be transcribed into an RNA molecule,
10 such as a messenger RNA (mRNA) molecule. Expression may further include translation of an RNA molecule and translated into peptides, polypeptides, or proteins. If the nucleic acid is derived from genomic DNA, expression may, if an appropriate eukaryotic host cell or organism is selected, include splicing of the mRNA. With respect to an antisense
15 construct, expression may refer to the transcription of the antisense DNA.

As used herein, vector or plasmid refers to discrete elements that are used to introduce heterologous nucleic acids into cells for either expression of the heterologous nucleic acid or for replication of the heterologous nucleic acid. Selection and use of such vectors and
20 plasmids are well within the level of skill of the art.

As used herein, transformation/transfection refers to the process by which nucleic acid is introduced into cells. The terms transfection and transformation refer to the taking up of exogenous nucleic acid, e.g., an expression vector, by a host cell whether or not any coding sequences
25 are in fact expressed. Numerous methods of transfection are known to the ordinarily skilled artisan, for example, by *Agrobacterium*-mediated transformation, protoplast transformation (including polyethylene glycol (PEG)-mediated transformation, electroporation, protoplast fusion, and microcell fusion), lipid-mediated delivery, liposomes, electroporation,

sonoporation, microinjection, particle bombardment and silicon carbide whisker-mediated transformation and combinations thereof (see, *e.g.*, Paszkowski *et al.* (1984) *EMBO J.* 3:2717-2722; Potrykus *et al.* (1985) *Mol. Gen. Genet.* 199:169-177; Reich *et al.* (1986) *Biotechnology* 4:1001-1004; Klein *et al.* (1987) *Nature* 327:70-73; U.S. Patent No. 6,143,949; Paszkowski *et al.* (1989) in *Cell Culture and Somatic Cell Genetics of Plants, Vol. 6*, Molecular Biology of Plant Nuclear Genes, eds. Schell, J and Vasil, L.K. Academic Publishers, San Diego, California, p. 52-68; and Frame *et al.* (1994) *Plant J.* 6:941-948), direct uptake using calcium phosphate (CaPO₄; see, *e.g.*, Wigler *et al.* (1979) *Proc. Natl. Acad. Sci. U.S.A.* 76:1373-1376), polyethylene glycol (PEG)-mediated DNA uptake, lipofection (see, *e.g.*, Strauss (1996) *Meth. Mol. Biol.* 54:307-327), microcell fusion (see, EXAMPLES, see, also Lambert (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88:5907-5911; U.S. Patent No. 5,396,767, Sawford *et al.* (1987) *Somatic Cell Mol. Genet.* 13:279-284; Dhar *et al.* (1984) *Somatic Cell Mol. Genet.* 10:547-559; and McNeill-Killary *et al.* (1995) *Meth. Enzymol.* 254:133-152), lipid-mediated carrier systems (see, *e.g.*, Teifel *et al.* (1995) *Biotechniques* 19:79-80; Albrecht *et al.* (1996) *Ann. Hematol.* 72:73-79; Holmen *et al.* (1995) *In Vitro Cell Dev. Biol. Anim.* 31:347-351; Remy *et al.* (1994) *Bioconjug. Chem.* 5:647-654; Le Bolch *et al.* (1995) *Tetrahedron Lett.* 36:6681-6684; Loeffler *et al.* (1993) *Meth. Enzymol.* 217:599-618) or other suitable method. Methods for delivery of ACes are described in copending U.S. application Serial No. 09/815,979. Successful transfection is generally recognized by detection of the presence of the heterologous nucleic acid within the transfected cell, such as, for example, any visualization of the heterologous nucleic acid or any indication of the operation of a vector within the host cell.

As used herein, "delivery," which is used interchangeably with "transfection," refers to the process by which exogenous nucleic acid molecules are transferred into a cell such that they are located inside the cell. Delivery of nucleic acids is a distinct process from expression of
5 nucleic acids.

As used herein, injected refers to the microinjection, such as by use of a small syringe, needle, or pipette, for injection of nucleic acid into a cell.

As used herein, substantially homologous DNA refers to DNA that
10 includes a sequence of nucleotides that is sufficiently similar to another such sequence to form stable hybrids, with each other or a reference sequence, under specified conditions.

It is well known to those of skill in this art that nucleic acid fragments with different sequences may, under the same conditions,
15 hybridize detectably to the same "target" nucleic acid. Two nucleic acid fragments hybridize detectably, under stringent conditions over a sufficiently long hybridization period, because one fragment contains a segment of at least about 10, 14 or 16 or more nucleotides in a sequence that is complementary (or nearly complementary) to a substantially
20 contiguous sequence of at least one segment in the other nucleic acid fragment. If the time during which hybridization is allowed to occur is held constant, at a value during which, under preselected stringency conditions, two nucleic acid fragments with complementary base-pairing segments hybridize detectably to each other, departures from exact
25 complementarity can be introduced into the base-pairing segments, and base-pairing will nonetheless occur to an extent sufficient to make hybridization detectable. As the departure from complementarity between the base-pairing segments of two nucleic acids becomes larger, and as

conditions of the hybridization become more stringent, the probability decreases that the two segments will hybridize detectably to each other.

Two single-stranded nucleic acid segments have "substantially the same sequence", if (a) both form a base-paired duplex with the same
5 segment, and (b) the melting temperatures of the two duplexes in a solution of 0.5 X SSPE differ by less than 10°C. If the segments being compared have the same number of bases, then to have "substantially the same sequence", they will typically differ in their sequences at fewer than 1 base in 10. Methods for determining melting temperatures of
10 nucleic acid duplexes are well known (see, *e.g.*, Meinkoth *et al.* (1984) *Anal. Biochem.* 138:267-284 and references cited therein).

As used herein, a nucleic acid probe is a DNA or RNA fragment that includes a sufficient number of nucleotides to specifically hybridize to DNA or RNA that includes complementary or substantially complementary
15 sequences of nucleotides. A probe may contain any number of nucleotides, from as few as about 10 and as many as hundreds of thousands of nucleotides. The conditions and protocols for such hybridization reactions are well known to those of skill in the art as are the effects of probe size, temperature, degree of mismatch, salt
20 concentration and other parameters on the hybridization reaction. For example, the lower the temperature and higher the salt concentration at which the hybridization reaction is carried out, the greater the degree of mismatch that may be present in the hybrid molecules.

To be used as a hybridization probe, the nucleic acid is generally
25 rendered detectable by labeling it with a detectable moiety or label, such as ³²P, ³H and ¹⁴C, or by other means, including chemical labeling, such as by nick-translation in the presence of deoxyuridylate biotinylated at the 5'-position of the uracil moiety. The resulting probe includes the biotinylated uridylate in place of thymidylate residues and can be detected

(via the biotin moieties) by any of a number of commercially available detection systems based on binding of streptavidin to the biotin. Such commercially available detection systems can be obtained, for example, from Enzo Biochemicals, Inc. (New York, NY). Any other label known to
5 those of skill in the art, including non-radioactive labels, may be used as long as it renders the probes sufficiently detectable, which is a function of the sensitivity of the assay, the time available (for culturing cells, extracting DNA, and hybridization assays), the quantity of DNA or RNA available as a source of the probe, the particular label and the means used
10 to detect the label.

Once sequences with a sufficiently high degree of homology to the probe are identified, they can readily be isolated by standard techniques (see, *e.g.*, Sambrook *et al.* (2001) *Molecular Cloning: A Laboratory Manual*, 3rd Edition, Cold Spring Harbor Laboratory Press).

15 As used herein, conditions under which DNA molecules form stable hybrids are considered substantially homologous, and a DNA or nucleic acid homolog refers to a nucleic acid that includes a preselected conserved nucleotide sequence, such as a sequence encoding a polypeptide. By the term "substantially homologous" is meant having at
20 least 75%, preferably 80%, preferably at least 90%, most preferably at least 95% homology therewith or a less percentage of homology or identity and conserved biological activity or function.

The terms "homology" and "identity" are often used interchangeably. In this regard, percent homology or identity may be
25 determined, for example, by comparing sequence information using a GAP computer program. The GAP program utilizes the alignment method of Needleman and Wunsch (*J. Mol. Biol.* 48:443 (1970), as revised by Smith and Waterman (*Adv. Appl. Math.* 2:482 (1981)). Briefly, the GAP program defines similarity as the number of aligned symbols (i.e.,

nucleotides or amino acids) which are similar, divided by the total number of symbols in the shorter of the two sequences. The preferred default parameters for the GAP program may include: (1) a unary comparison matrix (containing a value of 1 for identities and 0 for non-identities) and
5 the weighted comparison matrix of Gribskov and Burgess, *Nucl. Acids Res.* 14:6745 (1986), as described by Schwartz and Dayhoff, eds., *ATLAS OF PROTEIN SEQUENCE AND STRUCTURE*, National Biomedical Research Foundation, pp. 353-358 (1979); (2) a penalty of 3.0 for each gap and an additional 0.10 penalty for each symbol in each gap; and (3)
10 no penalty for end gaps.

By sequence identity, the number of conserved amino acids are determined by standard alignment algorithms programs, and are used with default gap penalties established by each supplier. Substantially homologous nucleic acid molecules would hybridize typically at moderate
15 stringency or at high stringency all along the length of the nucleic acid of interest. Preferably the two molecules will hybridize under conditions of high stringency. Also contemplated are nucleic acid molecules that contain degenerate codons in place of codons in the hybridizing nucleic acid molecule.

20 Whether any two nucleic acid molecules have nucleotide sequences that are at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% "identical" can be determined using known computer algorithms such as the "FAST A" program, using for example, the default parameters as in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 (1988).
25 Alternatively the BLAST function of the National Center for Biotechnology Information database may be used to determine relative sequence identity.

In general, sequences are aligned so that the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be

calculated using published techniques. (See, e.g.: *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 5 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or 10 polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H. & Lipton, D., *SIAM J Applied Math* 48:1073 (1988)). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 15 1994, and Carillo, H. & Lipton, D., *SIAM J Applied Math* 48:1073 (1988). Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCG program package (Devereux, J., et al., *Nucleic Acids Research* 20 12(II):387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F., et al., *J Molec Biol* 215:403 (1990)).

Therefore, as used herein, the term "identity" represents a comparison between a test and a reference polypeptide or polynucleotide. For example, a test polypeptide may be defined as any polypeptide that is 25 90% or more identical to a reference polypeptide.

As used herein, the term at least "90% identical to" refers to percent identities from 90 to 99.99 relative to the reference polypeptides. Identity at a level of 90% or more is indicative of the fact that, assuming for exemplification purposes a test and reference polynucleotide length of

100 amino acids are compared. No more than 10% (i.e., 10 out of 100) amino acids in the test polypeptide differs from that of the reference polypeptides. Similar comparisons may be made between a test and reference polynucleotides. Such differences may be represented as point mutations randomly distributed over the entire length of an amino acid sequence or they may be clustered in one or more locations of varying length up to the maximum allowable, e.g. 10/100 amino acid difference (approximately 90% identity). Differences are defined as nucleic acid or amino acid substitutions, or deletions.

10 As used herein: stringency of hybridization in determining percentage mismatch encompass the following conditions or equivalent conditions thereto:

- 1) high stringency: 0.1 x SSPE or SSC, 0.1% SDS, 65°C
- 2) medium stringency: 0.2 x SSPE or SSC, 0.1% SDS, 50°C
- 15 3) low stringency: 1.0 x SSPE or SSC, 0.1% SDS, 50°C

or any combination of salt and temperature and other reagents that result in selection of the same degree of mismatch or matching. Equivalent conditions refer to conditions that select for substantially the same percentage of mismatch in the resulting hybrids. Additions of ingredients, such as formamide, Ficoll, and Denhardt's solution affect parameters such as the temperature under which the hybridization should be conducted and the rate of the reaction. Thus, hybridization in 5 X SSC, in 20% formamide at 42° C is substantially the same as the conditions recited above hybridization under conditions of low stringency. The recipes for SSPE, SSC and Denhardt's and the preparation of deionized formamide are described, for example, in Sambrook *et al.* (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Chapter 8; see, Sambrook *et al.*, vol. 3, p. B.13, see, also, numerous catalogs that describe commonly used laboratory solutions. It is understood that

equivalent stringencies may be achieved using alternative buffers, salts and temperatures. As used herein, all assays and procedures, such as hybridization reactions and antibody-antigen reactions, unless otherwise specified, are conducted under conditions recognized by those of skill in the art as standard conditions.

As used herein, conservative amino acid substitutions, such as those set forth in Table 1, are those that do not eliminate biological activity. Suitable conservative substitutions of amino acids are known to those of skill in this art and may be made generally without altering the biological activity of the resulting molecule. Those of skill in this art recognize that, in general, single amino acid substitutions in non-essential regions of a polypeptide do not substantially alter biological activity (see, *e.g.*, Watson *et al. Molecular Biology of the Gene*, 4th Edition, 1987, The Bejacmin/Cummings Pub. co., p.224). Conservative amino acid substitutions are made, for example, in accordance with those set forth in TABLE 1 as follows:

TABLE 1

	Original residue	Conservative substitution
	Ala (A)	Gly; Ser, Abu
20	Arg (R)	Lys, orn
	Asn (N)	Gln; His
	Cys (C)	Ser
	Gln (Q)	Asn
	Glu (E)	Asp
25	Gly (G)	Ala; Pro
	His (H)	Asn; Gln
	Ile (I)	Leu; Val; Met; Nle; Nva
	Leu (L)	Ile; Val; Met; Nle; Nva
	Lys (K)	Arg; Gln; Glu
30	Met (M)	Leu; Tyr; Ile; NLe Val
	Ornithine	Lys; Arg
	Phe (F)	Met; Leu; Tyr
	Ser (S)	Thr
	Thr (T)	Ser
35	Trp (W)	Tyr
	Tyr (Y)	Trp; Phe
	Val (V)	Ile; Leu; Met; Nle; Nva

Other substitutions are also permissible and may be determined empirically or in accord with known conservative substitutions.

As used herein, the amino acids, which occur in the various amino acid sequences appearing herein, are identified according to their well-known, three-letter or one-letter abbreviations. The nucleotides, which occur in the various DNA fragments, are designated with the standard single-letter designations used routinely in the art.

As used herein, a splice variant refers to a variant produced by differential processing of a primary transcript of genomic DNA that results in more than one type of mRNA.

As used herein, a probe or primer based on a nucleotide sequence includes at least 10, 14, 16, 30 or 100 contiguous nucleotides from the reference nucleic acid molecule.

As used herein, recombinant production by using recombinant DNA methods refers to the use of the well known methods of molecular biology for expressing proteins encoded by cloned DNA.

As used herein, biological activity refers to the *in vivo* activities of a compound or physiological responses that result upon *in vivo* administration of a compound, composition or other mixture. Biological activity, thus, encompasses therapeutic effects and pharmaceutical activity of such compounds, compositions and mixtures. Biological activities may be observed in *in vitro* systems designed to test or use such activities. Thus, for purposes herein the biological activity of a luciferase is its oxygenase activity whereby, upon oxidation of a substrate, light is produced.

The terms substantially identical or similar varies with the context as understood by those skilled in the relevant art and generally means at least 40, 60, 80, 90, 95 or 98%.

As used herein, substantially identical to a product means
5 sufficiently similar so that the property is sufficiently unchanged so that the substantially identical product can be used in place of the product.

As used herein, substantially pure means sufficiently homogeneous to appear free of readily detectable impurities as determined by standard methods of analysis, such as thin layer chromatography (TLC), gel
10 electrophoresis and high performance liquid chromatography (HPLC), used by those of skill in the art to assess such purity, or sufficiently pure such that further purification would not detectably alter the physical and chemical properties, such as enzymatic and biological activities, of the substance. Methods for purification of the compounds to produce
15 substantially chemically pure compounds are known to those of skill in the art. A substantially chemically pure compound may, however, be a mixture of stereoisomers or isomers. In such instances, further purification might increase the specific activity of the compound.

As used herein, vector (or plasmid) refers to discrete elements that
20 are used to introduce heterologous DNA into cells for either expression or replication thereof. The vectors typically remain episomal, but may be designed to effect integration of a gene or portion thereof into a chromosome of the genome. Also contemplated are vectors that are artificial chromosomes, such as yeast artificial chromosomes and
25 mammalian artificial chromosomes. Selection and use of such vehicles are well known to those of skill in the art. An expression vector includes vectors capable of expressing DNA that is operatively linked with regulatory sequences, such as promoter regions, that are capable of effecting expression of such DNA fragments. Thus, an expression vector

refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector that, upon introduction into an appropriate host cell, results in expression of the cloned DNA.

Appropriate expression vectors are well known to those of skill in the art and include those that are replicable in eukaryotic cells and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome.

As used herein, protein-binding-sequence refers to a protein or peptide sequence that is capable of specific binding to other protein or peptide sequences generally, to a set of protein or peptide sequences or to a particular protein or peptide sequence.

As used herein, a composition refers to any mixture of two or more ingredients. It may be a solution, a suspension, liquid, powder, a paste, aqueous, non-aqueous or any combination thereof.

As used herein, a combination refers to any association between two or more items.

As used herein, fluid refers to any composition that can flow. Fluids thus encompass compositions that are in the form of semi-solids, pastes, solutions, aqueous mixtures, gels, lotions, creams and other such compositions.

As used herein, a cellular extract refers to a preparation or fraction that is made from a lysed or disrupted cell.

As used herein, the term "subject" refers to animals, plants, insects, and birds and other phyla, genera and species into which nucleic acid molecules may be introduced. Included are higher organisms, such as mammals, fish, insects and birds, including humans, primates, cattle, pigs, rabbits, goats, sheep, mice, rats, guinea pigs, hamsters, cats, dogs, horses, chicken and others.

As used herein, flow cytometry refers to processes that use a laser based instrument capable of analyzing and sorting out cells and or chromosomes based on size and fluorescence.

As used herein, the abbreviations for any protective groups, amino
5 acids and other compounds, are, unless indicated otherwise, in accord with their common usage, recognized abbreviations, or the IUPAC-IUB Commission on Biochemical Nomenclature (see, (1972) *Biochem.* 11:942-944).

B. Recombination systems

10 Site-specific recombination systems typically contain three elements: a pair of DNA sequences (the site-specific recombination sequences) and a specific enzyme (the site-specific recombinase). The site-specific recombinase catalyzes a recombination reaction between two site-specific recombination sequences.

15 A number of different site-specific recombinase systems are available and/or known to those of skill in the art, including, but not limited to: the Cre/*lox* recombination system using CRE recombinase (see, *e.g.*, SEQ ID Nos. 58 and 59) from the Escherichia coli phage P1 (see, *e.g.*, Sauer (1993) *Methods in Enzymology* 225:890-900; Sauer *et al.* 20 (1990) *The New Biologist* 2:441-449), Sauer (1994) *Current Opinion in Biotechnology* 5:521-527; Odell *et al.* (1990) *Mol Gen Genet.* 223:369-378; Lasko *et al.* (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:6232-6236; U.S. Patent No. 5,658,772), the FLP/FRT system of yeast using the FLP recombinase (see, SEQ ID Nos. 60 and 61) from the 2 μ episome of 25 *Saccharomyces cerevisiae* (Cox (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80:4223; Falco *et al.* (1982) *Cell* 29:573-584; Golic *et al.* (1989) *Cell* 59:499-509; U.S. Patent No. 5,744,336), the resolvases, including Gin recombinase of phage Mu (Maeser *et al.* (1991) *Mol Gen Genet.* 230:170-176; Klippel, A. *et al* (1993) *EMBO J.* 12:1047-1057; see, *e.g.*,

SEQ ID Nos. 64-67), Cin, Hin, $\alpha\delta$ Tn3; the Pin recombinase of *E. coli* (see, e.g., SEQ ID Nos. 68 and 69; Enomoto *et al.* (1983) *J. Bacteriol.* 6:663-668), the R/RS system of the pSR1 plasmid of *Zygosaccharomyces rouxii* (Araki *et al.* (1992) *J. Mol. Biol.* 225:25-37; Matsuzaki *et al.* (1990) *J. Bacteriol.* 172: 610-618) and site-specific recombinases from *Kluyveromyces drosophilarius* (Chen *et al.* (1986) *Nucleic Acids Res.* 314:4471-4481) and *Kluyveromyces waltii* (Chen *et al.* (1992) *J. Gen. Microbiol.* 138:337-345). Other systems are known to those of skill in the art (Stark *et al.* *Trends Genet.* 8:432-439; Utatsu *et al.* (1987) *J. Bacteriol.* 169:5537-5545; see, also, U.S. Patent No. 6,171,861).

Members of the highly related family of site-specific recombinases, the resolvase family, such as $\gamma\delta$, Tn3 resolvase, Hin, Gin, and Cin are also available. Members of this family of recombinases are typically constrained to intramolecular reactions (e.g., inversions and excisions) and can require host-encoded factors. Mutants have been isolated that relieve some of the requirements for host factors (Maeser *et al.* (1991) *Mol. Gen. Genet.* 230:170-176), as well as some of the constraints of intramolecular recombination (see, U.S. Patent No. 6,171,861).

The bacteriophage P1 Cre/lox and the yeast FLP/FRT systems are particularly useful systems for site-specific integration, inversion or excision of heterologous nucleic acid into, and out of, chromosomes, particularly ACes as provided herein. In these systems a recombinase (Cre or FLP) interacts specifically with its respective site-specific recombination sequence (lox or FRT, respectively) to invert or excise the intervening sequences. The sequence for each of these two systems is relatively short (34 bp for lox and 47 bp for FRT).

The FLP/FRT recombinase system has been demonstrated to function efficiently in plant cells (U.S. Patent No. 5,744,386), and, thus, can be used for producing plant artificial chromosome platforms. In

general, short incomplete FRT sites leads to higher accumulation of excision products than the complete full-length FRT sites. The system catalyzes intra- and intermolecular reactions, and, thus, can be used for DNA excision and integration reactions. The recombination reaction is
5 reversible and this reversibility can compromise the efficiency of the reaction in each direction. Altering the structure of the site-specific recombination sequences is one approach to remedying this situation. The site-specific recombination sequence can be mutated in a manner that the product of the recombination reaction is no longer recognized as
10 a substrate for the reverse reaction, thereby stabilizing the integration or excision event.

In the Cre-lox system, discovered in bacteriophage P1, recombination between loxP sites occurs in the presence of the Cre recombinase (see, *e.g.*, U.S. Patent No. 5,658,772). This system can be
15 used to insert, invert or excise nucleic acid located between two lox sites. Cre can be expressed from a vector. Since the lox site is an asymmetrical nucleotide sequence, lox sites on the same DNA molecule can have the same or opposite orientation with respect to each other. Recombination between lox sites in the same orientation results in a deletion of the DNA
20 segment located between the two lox sites and a connection between the resulting ends of the original DNA molecule. The deleted DNA segment forms a circular molecule of DNA. The original DNA molecule and the resulting circular molecule each contain a single lox site. Recombination between lox sites in opposite orientations on the same DNA molecule
25 result in an inversion of the nucleotide sequence of the DNA segment located between the two lox sites. In addition, reciprocal exchange of DNA segments proximate to lox sites located on two different DNA molecules can occur. All of these recombination events are catalyzed by the product of the Cre coding region.

Any site-specific recombinase system known to those of skill in the art is contemplated for use herein. It is contemplated that one or a plurality of sites that direct the recombination by the recombinase are introduced into an artificial chromosome to produce platform *ACes*. The
5 resulting platform *ACes* are introduced into cells with nucleic acid encoding the cognate recombinase, typically on a vector, and nucleic acid encoding heterologous nucleic acid of interest linked to the appropriate recombination site for insertion into the platform *ACes*. The recombinase-encoding-nucleic acid may be introduced into the cells on the same
10 vector, or a different vector, encoding the heterologous nucleic acid.

An *E. coli* phage lambda integrase system for *ACes* platform engineering and for artificial chromosome engineering is provided (Lorbach *et al.* (2000) *J. Mol. Biol* 296:1175-1181). The phage lambda integrase (Landy, A. (1989) *Annu. Rev. Biochem.* 58:913-94) is adapted herein and
15 the cognate *att* sites are provided. Chromosomes, including *ACes*, engineered to contain one or a plurality of *att* sites are provided, as are vectors encoding a mutant integrase that functions in the absence other factors. Methods using the modified chromosomes and vectors for introduction of heterologous nucleic acid are also provided.

20 For purposes herein, one or more of the sites (e.g., a single site or a pair of sites) required for recombination are introduced into an artificial chromosome, such as an *ACes* chromosome. The enzyme for catalyzing site-directed recombination is introduced with the DNA of interest, or separately, or is engineered onto the artificial chromosome under the
25 control of a regulatable promoter.

As described herein, artificial chromosome platforms containing one or multiple recombination sites are provided. The methods and resulting products are exemplified with the lambda phage *Att/Int* system, but

similar methods may be used for production of *ACes* platforms with other recombination systems.

The *Att/Int* system and vectors provided herein are not only intended for engineering *ACes* platforms, but may be used to engineer an
5 *Att/Int* system into any chromosome. Introduction of *att* sites into a chromosome will permit engineering of natural chromosomes, such as by permitting targeted integration genes or regulatory regions, and by controlled excision of selected regions. For example, genes encoding a particular trait may be added to a chromosome, such as plant
10 chromosome engineered to contain one or plurality of *att* sites. Such chromosomes may be used for screening DNA to identify genes. Large pieces of DNA can be introduced into cells and the cells screened phenotypically to select those having the desired trait.

C. Platforms

15 Provided herein are platform artificial chromosomes (platform *ACes*) containing single or multiple site-specific recombination sites. Chromosome-based platform technology permits efficient and tractable engineering and subsequent expression of multiple gene targets. Methods are provided that use DNA vectors and fragments to create platform
20 artificial chromosomes, including animal, particularly mammalian, artificial chromosomes, and plant artificial chromosomes. The artificial chromosomes contain either single or multiple sequence-specific recombination sites suitable for the placement of target gene expression vectors onto the platform chromosome. The engineered chromosome-
25 based platform *ACes* technology is applicable for methods, including cellular and transgenic protein production, transgenic plant and animal production and gene therapy. The platform *ACes* are also useful for producing a library of *ACes* comprising random portions of a given genome (e.g., a mammalian, plant or prokaryotic genome) for genomic

screening; as well as a library of cells comprising different and/or mutually exclusive *ACes* therein.

Exemplary of artificial chromosome platforms are those based on *ACes*. *ACes* artificial chromosomes are non-viral, self-replicating nucleic acid molecules that function as a natural chromosome, having all the elements required for normal chromosomal replication and maintenance within the cell nucleus. *ACes* artificial chromosomes do not rely on integration into the genome of the cell to be effective, and they are not limited by DNA carrying capacity and as such the therapeutic gene(s) of interest, including regulatory sequences, can be engineered into the *ACes*. In addition, *ACes* are stable *in vitro* and *in vivo* and can provide predictable long-term gene expression. Once engineered and delivered to the appropriate cell or embryo, *ACes* work independently alongside host chromosomes, for *ACes* that are predominantly heterochromatin producing only the products (proteins) from the genes it carries. As provided herein *ACes* are modified by introduction of recombination site(s) to provide a platform for ready introduction of heterologous nucleic acid. The *ACes* platforms can be used for production of transgenic animals and plants; as vectors for genetic therapy; for use as protein production systems; for animal models to identify and target new therapeutics; in cell culture for the development and production of therapeutic proteins; and for a variety of other applications.

1. Generation of artificial chromosomes

Artificial chromosomes may be generated by any method known to those of skill in the art. Of particular interest herein are the *ACes* artificial chromosomes, which contain a repeated unit. Methods for production of *ACes* are described in detail in U.S. Patent Nos. 6,025,155 and 6,077,697, which, as with all patents, applications, publications and other disclosure, are incorporated herein in their entirety.

Generation of *de novo* ACes.

ACes can be generated by cotransfecting exogenous DNA—such as a mammary tissue specific DNA cassette including the gene sequences for a therapeutic protein, with a rDNA fragment and a drug resistance
5 marker gene into the desired eukaryotic cell, such as plant or animal cells, such as murine cells *in vitro*. DNA with a selectable or detectable marker is introduced, and can be allowed to integrate randomly into pericentric heterochromatin or can be targeted to pericentric heterochromatin, such as that in rDNA gene arrays that reside on acrocentric chromosomes,
10 such as the short arms of acrocentric chromosomes. This integration event activates the “megareplicator” sequence and amplifies the pericentric heterochromatin and the exogenous DNA, and duplicates a centromere. Ensuing breakage of this “dicentric” chromosome can result in the production of daughter cells that contain the substantially-original
15 chromosome and the new artificial chromosome. The resulting ACes contain all the essential elements needed for stability and replication in dividing cells—centromere, origins of replications, and telomeres. ACes have been produced that express marker genes (lacZ, green fluorescent protein, neomycin-resistance, puromycin-resistance, hygromycin-
20 resistance) and genes of interest. Isolated ACes, for example, have been successfully transferred intact to rodent, human, and bovine cells by electroporation, sonoporation, microinjection, and transfection with lipids and dendrimers.

To render the creation of ACes with desired genes more tractable
25 and efficient, “platform” ACes (platform-ACes) can be produced that contain defined DNA sequences for enzyme-mediated homologous DNA recombination, such as by Cre or FLP recombinases (Bouhassira *et al.* (1996) *Blood* 88(supplement 1):190a; Bouhassira *et al.* (1997) *Blood*, 90:3332-3344; Siebler *et al.* (1997) *Biochemistry*: 36:1740-1747;

Siebler *et al.* (1998) *Biochemistry* 37: 6229-6234; and Bethke *et al.* (1997) *Nucl. Acids Res.* 25:2828-2834), and as exemplified herein the lambda phage integrase. A *lox* site contains two 13 bp inverted repeats to which Cre-recombinase binds and an intervening 8 bp core region.

- 5 Only pairs of sites having identity in the central 6 bp of the core region are proficient for recombination; sites having non-identical core sequences (heterospecific *lox* sites) do not efficiently recombine with each other (Hoess *et al.* (1986) *Nucleic Acids Res.* 14:2287-2300).

10 **Generating acrocentric chromosomes for plant artificial chromosome formation.**

In human and mouse cells *de novo* formation of a satellite DNA based artificial chromosome (SATAC, also referred to as ACes) can occur in an acrocentric chromosome where the short arm contains only pericentric heterochromatin, the rDNA array, and telomere sequences.

- 15 Plant species may not have any acrocentric chromosomes with the same physical structure described, but "megareplicator" DNA sequences reside in the plant rDNA arrays, also known as the nucleolar organizing regions (NOR). A structure like those seen in acrocentric mammalian chromosomes can be generated using site-specific recombination between
20 appropriate arms of plant chromosomes.

Approach

- Qin *et al.* ((1994) *Proc. Natl. Acad. Sci. U.S.A.* 91:1706-1710, 1994) describes crossing two *Nicotiana tabacum* transgenic plants. One plant contains a construct encoding a promoterless hygromycin-resistance
25 gene preceded by a *lox* site (*lox-hpt*), the other plant carries a construct containing a cauliflower mosaic virus 35S promoter linked to a *lox* sequence and the *cre* DNA recombinase coding region (35S-*lox-cre*). The constructs were introduced separately by infecting leaf explants with *agrobacterium tumefaciens* which carries the kanamycin-resistance gene

(Kan^R). The resultant Kan^R transgenic plants were crossed. Plants that carried the appropriate DNA recombination event were identified by hygromycin-resistance.

5 **Modification of the above for generation of ACes**

 The Kan^R cultivars are initially screened, such as by FISH, to identify two sets of candidate transgenic plants. One set has one construct integrated in regions adjacent to the pericentric heterochromatin on the short arm of any chromosome. The second set of candidate plants
10 has the other construct integrated in the NOR region of appropriate chromosomes. To obtain reciprocal translocation both sites must be in the same orientation. Therefore a series of crosses are required, Kan^R plants generated, and FISH analyses performed to identify the appropriate "acrocentric" plant chromosome for *de novo* plant ACes formation.

15 **2. Bacteriophage lambda integrase-based site-specific recombination system**

 An integral part of the platform technology includes a site-specific recombination system that allows the placement of selected gene targets or genomic fragments onto the platform chromosomes. Any such system
20 may be used. In particular, a method is provided for insertion of additional DNA fragments into the platform chromosome residing in the cell via sequence-specific recombination using the recombinase activity of the bacteriophage lambda integrase. The lambda integrase system is exemplary of the recombination systems contemplated for ACes. Any
25 known recombination system, including any described herein, particularly any that operates without the need for additional factors or that, by virtue of mutation, does not require additional factors, is contemplated.

As noted the lambda integrase system provided herein can be used with natural chromosomes and artificial chromosomes in addition to ACes. Single or a plurality of recombination sites, which may be the same or different, are introduced into artificial chromosomes to produce artificial chromosome platforms.

3. Creation of bacteriophage lambda integrase site-specific recombination system

The lambda phage-encoded integrase (designated Int) is a prototypical member of the integrase family. Int effects integration and excision of the phage in and out of the *E. coli* genome via recombination between pairs of attachment sites designated *attB/attP* and *attL/attR*. Each *att* site contains two inverted 9 base pair core Int binding sites and a 7 base pair overlap region that is identical in wild-type *att* sites. Each site, except for *attB* contains additional Int binding sites. In flanking regions, there are recognition sequences for accessory DNA binding proteins, such as integration host factor (IHF), factor for inversion stimulation (FIS) and the phage encoded excision protein (XIS). Except for *attB*, Int is a heterobivalent DNA-binding protein and, with assistance from the accessory proteins and negative DNA supercoiling, binds simultaneously to core and arm sites within the same *att* site.

Int, like Cre and FLP, executes an ordered sequential pair of strand exchanges during integrative and excisive recombination. The natural pairs of target sequences for Int, *attB* and *attP* or *attL* and *attR* are located on the same or different DNA molecules resulting in intra or intermolecular recombination, respectively. For example, intramolecular recombination occurs between inversely oriented *attB* and *attP*, or between *attL* and *attR* sequences, respectively, leading to inversion of the intervening DNA segment.

Like the recombinase systems, such as Cre and FLP, Int directs site-specific recombination. Unlike the other systems, such as Cre and FLP, Int generally requires additional protein factors for integrative and excisive recombination and negative supercoiling for integrative recombination.

5 Hence, the Int system had not been used in eukaryotic targeting systems.

Mutant Int proteins, designated Int-h (E174K) and a derivative thereof Int-h/218(E174K/E218K) do not require accessory proteins to perform intramolecular integrative and excisive recombination in co-transfection assays in human cells (Lorbach *et al.* (2000) *J Mol. Biol.*
10 296:1175-1181); wild-type Int does not catalyze intramolecular recombination in human cells harboring target sites *attB* and *attP*. Hence it had been demonstrated that mutant Int can catalyze factor-independent recombination events in human cells.

There has been no demonstration by others that this system can be
15 used for engineering of eukaryotic genomes or chromosomes. Provided herein are chromosomes, including artificial chromosomes, such as but not limited to *ACes* that contain *att* sites (e.g., platform *ACes*), and the use of such chromosomes for targeted integration of heterologous DNA into such chromosomes in eukaryotic cells, including animal, such as
20 rodent and human, and plant cells. Mutant Int provided herein is shown to effect site-directed recombination between sites in artificial chromosomes and vectors containing cognate sites.

An additional component of the chromosome-based platform technology is the site-specific integration of target DNA sequences onto
25 the platform. For this the native bacteriophage lambda integrase has been modified to carry out this sequence specific DNA recombination event in eukaryotic cells. The bacteriophage lambda integrase and its cognate DNA substrate *att* is a member of the site-specific recombinase family that also includes the bacteriophage P1 Cre/lox system as well as

the *Saccharomyces cerevisiae* 2 micron based FLP/FRT system (see, e.g., Landy (1989) *Ann. Rev. Biochem* 58:913-949; Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398-3402; Broach *et al.* (1982) *Cell* 29:227-234).

5 By combining DNA endonuclease and DNA ligase activity these recombinases recognize and catalyze DNA exchanges between sequences flanking the recognition site. During the integration of lambda genome into the *E. coli* (lambda recombination) genome, the phage integrase (INT) in association with accessory proteins catalyzes the DNA exchange
10 between the attP site of the phage genome and the attB site of the bacterial genome resulting in the formation of attL and attR sites (Figure 6). The engineered bacteriophage lambda integrase has been produced herein to carry out an intermolecular DNA recombination event between
15 an incoming DNA molecule (primarily on a vector containing the bacterial attB site) and the chromosome-based platform carrying the lambda attP sequence independent of lambda bacteriophage or bacterial accessory proteins.

 In contrast to the bi-directional Cre/lox and FLP/FRT system, the engineered lambda recombination system derived for chromosome-based
20 platform technology is advantageously unidirectional because accessory proteins, which are absent, are required for excision of integrated nucleic acid upon further exposure to the lambda Int recombinase.

4. Creation of platform chromosome containing single or multiple sequence-specific recombination sites

25 **a. Multiple sites**

 For the creation of a platform chromosome containing multiple, sequence-specific recombination sites, artificial chromosomes are produced as depicted in Figure 5 and Example 3. As discussed above, artificial chromosomes can be produced using any suitable methodology,

including those described in U.S. Patent Nos. 5,288,625; 5,712,134; 5,891,691; 6,025,155. Briefly, to prepare artificial chromosomes containing multiple recombination (e.g., integration) sites, nucleic acid (either in the form a one or more plasmids, such as the plasmid pSV40193attPsensePUR set forth in Example 3) is targeted into an amplifiable region of a chromosome, such as the pericentric region of a chromosome. Among such regions are the rDNA gene loci in acrocentric mammalian chromosomes. Hence, targeting nucleic acid for integration into the rDNA region of mammalian acrocentric chromosomes can include the mouse rDNA fragments (for targeting into rodent cell lines) or large human rDNA regions on BAC/PAC vectors (or subclones thereof in standard vectors) for targeting into human acrocentric chromosomes, such as for human gene therapy applications. The targeting nucleic acid generally includes a detectable or selectable marker, such as antibiotic resistance, such as puromycin and hygromycin, a recombination site (such as attP, attB, attL, attR or the like), and/or human selectable markers as required for gene therapy applications. Cells are grown under conditions that result in amplification and ultimately production of ACes artificial chromosomes having multiple recombination (e.g., integration) sites therein. ACes having the desired size are selected for further engineering.

b. Creation of platform chromosome containing a single sequence-specific recombination site

In this method a mammalian platform artificial chromosome is generated containing a single sequence-specific recombination site. In the Example below, this approach is demonstrated using a puromycin resistance marker for selection and a mouse rDNA fragment for targeting into the rDNA locus on mouse acrocentric chromosomes. Other selection markers and targeting DNA sequences as desired and known to those of

skill in the art can be used. Additional resistance markers include genes conferring resistance to the antibiotics neomycin, blasticidin, hygromycin and zeocin. For applications, such as gene therapy in which potentially immunogenic responses are to be avoided, host, such as human, derived

5 selectable markers or markers detectable with monoclonal antibodies (MAb) followed by fluorescent activated cell sorting (FACS) can be used. Examples in this class include, but are not limited to: human nerve growth factor receptor (detection with MAb); truncated human growth factor receptor (detection with MAb); mutant human dihydrofolate reductase

10 (DHFR; detectable using a fluorescent methotrexate substrate); secreted alkaline phosphatase (SEAP; detectable with fluorescent substrate); thymidylate synthase (TS; confers resistance to fluorodeoxyuridine); human CAD gene (confers resistance to N-phosphonacetyl-L-aspartate (PALA)).

15 To construct a platform artificial chromosome with a single site, an *ACes* artificial chromosome (or other artificial chromosome of interest) can be produced containing a selectable marker. A single sequence specific recombination site is targeted onto *ACes* via homologous recombination. For this, DNA sequences containing the site-specific

20 recombination sequence are flanked with DNA sequences homologous to a selected sequence in the chromosome. For example, when using a chromosome containing rDNA or satellite DNA, such DNA can be used as homologous sequences to target the site-specific recombination sequence onto the chromosome. A vector is designed to have these homologous

25 sequences flanking the site-specific recombination site and, after the appropriate restriction enzyme digest to generate free ends of homology to the chromosome, the DNA is transfected into cells harboring the chromosome. After transfection and integration of the site-specific cassette, homologous recombination events onto the platform

chromosome are subcloned and identified, for example by screening single cell subclones via expression of resistance or a fluorescent marker and PCR analysis. In one embodiment, a platform artificial chromosome, such as a platform ACes, that contains a single copy of the recombination site is selected. Examples 2B and 2D exemplify the process, and Figure 3 provides a diagram depicting one method for the creation of a platform mammalian chromosome containing a single sequence-specific recombination site.

10 5. **Lambda integrase mediated recombination of target gene expression vector onto platform chromosome**

 The third component of the chromosome-based platform technology involves the use of target gene expression vectors carrying, for example, genes for gene therapy, genes for transgenic animal or plant production, and those required for cellular protein production of interest. 15 Using lambda integrase mediated site-specific recombination, or any other recombinase-mediated site-specific recombination, the target gene expression vectors are introduced onto the selected chromosome platform. The use of target gene expression vector permits use of the *de novo* generated chromosome-based platforms for a wide range of gene 20 targets. Furthermore, chromosome platforms containing multiple *attP* sites provides the opportunity to incorporate multiple gene targets onto a single platform, thereby providing for expression of multiple gene targets, including the expression of cellular and genetic regulatory genes and the expression of all or parts of metabolic pathways. In addition to 25 expressing small target genes, such as cDNA and hybrid cDNA/artificial intron constructs, the chromosome-based platform can be used for engineering and expressing large genomic fragments carrying target genes along with its endogenous genomic promoter sequences. This is of importance, for example, where the therapy requires precise cell specific

expression and in instances where expression is best achieved from genomic clones rather than cDNA clones. Figure 9 provides a diagram summarizing one embodiment of the chromosome-based technology.

A feature of the target gene expression vector that is of interest to
5 include is a promoterless marker gene, which as exemplified (see, Figure 9) contains an upstream *attB* site (marker 2 on Figure 9). The nucleic acid encoding the marker is not expressed unless it is placed downstream from a promoter sequence. Using the recombinase technology provided herein, such as the lambda integrase technology (λ INT_{E174R} on figure 8)
10 provided herein, site-specific recombination between the *attB* site on the vector and the promoter-*attP* site (in the "sense" orientation) on the chromosome-based platform results in the expression of marker 2 on the target gene expression vector, thereby providing a positive selection for the lambda INT mediated site-specific recombination event. Site-specific
15 recombination events on the chromosome-based platform versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR. Examples of suitable marker 2 genes, include, but are not limited to, genes that confer resistance to toxic compounds or antibiotics,
20 fluorescence activated cell sorting (FACS) sortable cell surface markers and various fluorescent markers. Examples of these genes include, but are not limited to, human L26a^R (human homolog of *Saccharomyces cerevisiae* CYH^B gene), neomycin, puromycin, blasticidin, CD24 (see, e.g., US Patents 5,804,177 and 6,074,836), truncated CD4, truncated low
25 affinity nerve growth factor receptor (LNGFR), truncated LDL receptor, truncated human growth hormone receptor, GFP, RFP, BFP.

The target gene expression vectors contain a gene (target gene) for expression from the chromosome platform. The target gene can be expressed using various constitutive or regulated promoter systems

across various mammalian species. For the expression of multiple target genes within the same target gene expression vector, the expression of the multiple targets can be coordinately regulated via viral-based or human internal ribosome entry site (IRES) elements (see, *e.g.*, Jackson *et al.* (1990) *Trends Biochem Sci.* 15: 477-83; Oumard *et al.* (2000) *Mol. Cell. Biol.* 20: 2755-2759). Furthermore, using IRES type elements linked to a downstream fluorescent marker, *e.g.*, green, red or blue fluorescent proteins (GFP, RFP, BFP) allows for the identification of high expressing clones from the integrated target gene expression vector.

10 In certain embodiments described herein, the promoterless marker can be transcriptionally downstream of the heterologous nucleic acid, wherein the heterologous nucleic acid encodes a heterologous protein, and wherein the expression level of the selectable marker is transcriptionally linked to the expression level of the heterologous protein.

15 In addition, the selectable marker and the heterologous nucleic acid can be transcriptionally linked by the presence of a IRES between them. As set forth herein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic or fluorescent.

20 Expression from the target gene expression vector integrated onto the chromosome-based platform can be further enhanced using genomic insulator/boundary elements. The incorporation of insulator sequences into the target gene expression vector helps define boundaries in chromatin structure and thus minimizes influence of chromatin position

25 effects/gene silencing on the expression of the target gene (Bell *et al.* (1999) *Current Opinion in Genetics and Development* 9:191-198; Emery *et al.* (2000) *Proc. Natl. Acad. Sci. U.S.A.* 97:9150-9155). Examples of insulator elements that can be included onto target gene expression vector in order to optimize expression include, but are not limited to:

- 1) chicken β -globin HS4 element (Prioleau *et al.* (1999) *EMBO J* 18: 4035-4048);
- 2) matrix attachment regions (MAR; see, *e.g.*, Ramakrishnan *et al.* (2000) *Mol Cell. Biol.* 20:868-877);
- 5 3) scaffold attachment regions (SAR; see, *e.g.*, Auten *et al.* (1999) *Human Gene Therapy* 10:1389-1399); and
- 4) universal chromatin opening elements (UCOE; WO/0005393 and WO/0224930)

The copy number of the target gene can be controlled by sequentially adding multiple target gene expression vectors containing the target gene onto multiple integration sites on the chromosome platform. Likewise, the copy number of the target gene can be controlled within an individual target gene expression vector by the addition of DNA sequences that promote gene amplification. For example, gene amplification can be induced utilizing the dihydrofolate reductase (DHFR) minigene with subsequent selection with methotrexate (see, *e.g.*, Schimke (1984) *Cell* 37:705-713) or amplification promoting sequences from the rDNA locus (see, *e.g.*, Wegner *et al.* (1989) *Nucl. Acids Res.* 17: 9909-9932).

20 6. Platforms with other recombinase system sites

A "double *lox*" targeting strategy mediated by Cre-recombinase (Bethke *et al.* (1997) *Nucl. Acids Res.* 25:2828-2834) can be used. This strategy employs a pair of heterospecific *lox* sites—*loxA* and *loxB*, which differ by one nucleotide in the 8 bp spacer region. Both sites are engineered into the artificial chromosome and also onto the targeting DNA vector. This allows for a direct site-specific insertion of a commercially relevant gene or genes by a Cre-catalyzed double crossover event. In essence a platform ACes is engineered with a hygromycin-resistance gene flanked by the double *lox* sites generating *lox-ACes*, which is maintained

in the thymidine kinase deficient cell, LMtk(-). The gene of interest, for example, for testing purposes, the green fluorescence protein gene, GFP and a HSV thymidine kinase gene (tk) marker, are engineered between the appropriate *lox* sites of the targeting vector. The vector DNA is
5 cotransfected with plasmid pBS185 (Life Technologies) encoding the *Cre* recombinase gene into mammalian cells maintaining the dual-*lox* artificial chromosome. Transient expression of the *Cre* recombinase catalyzes the site-specific insertion of the gene and the tk-gene onto the artificial chromosome. The transfected cells are grown in HAT medium that
10 selects for only those cells that have integrated and expressed the thymidine kinase gene. The HAT^R colonies are screened by PCR analyses to identify artificial chromosomes with the desired insertion.

To generate the *lox-ACes*, Lambda-Hyg^R-*lox* DNA is transfected into the LMtk(-) cell line harboring the precursor *ACes*. Hygromycin-
15 resistant colonies are analyzed by FISH and Southern blotting for the presence of a single copy insert on the *ACes*.

To demonstrate the gene replacement technology, cell lines containing candidate *lox-ACes* are cotransfected with pTK-GFP-*lox* and pBS185 (encoding the *Cre* recombinase gene) DNA. After transfection,
20 transient expression of plasmid pBS185 will provide sufficient burst of *Cre* recombinase activity to catalyze DNA recombination at the *lox* sites. Thus, a double crossover event between the *ACes* target and the exogenous targeting plasmid carrying the *loxA* and *loxB* permits the simple replacement of the hygromycin-resistance gene on the *lox-ACes*
25 for the tk-GFP cassette from the targeting plasmid, with no integration of vector DNA. Transfected cells are grown in HAT-media to select for tk-expression. Correct targeting will result in the generation of HAT^R, hygromycin sensitive, and green fluorescent cells. The desired integration event is verified by Southern and PCR analyses. Specific PCR primer sets

are used to amplify DNA sequences flanking the individual *loxA* and *loxB* sites on the *lox-ACes* before and after homologous recombination.

D. Exemplary applications of the Platform *ACes*

Platform *ACes* are applicable and tractable for different/optimized
5 cell lines. Those that include a fluorescent marker, for example, can be
purified and isolated using fluorescent activated cell sorting (FACS), and
subsequently delivered to a target cell. Those with selectable markers
provide for efficient selection and provide a growth advantage. Platform
ACes allow multiple payload delivery of donor target vectors via a
10 positive-selection site-specific, recombination system, and they allow for
the inclusion of additional genetic factors that improve protein production
and protein quality.

The construction and use of the platform *ACes* as provided for
each application may be similarly applied to other applications. Particular
15 descriptions are for exemplification.

1. Cellular Protein Production Platform *ACes* (CPP *ACes*)

As described herein, *ACes* can be produced from acrocentric
chromosomes in rodent (mouse, hamster) cell lines via megareplicator
induced amplification of heterochromatin/rDNA sequences. Such *ACes*
20 are ideal for cellular protein production as well as other applications
described herein and known to those of skill in the art. *ACes* platforms
that contain a plurality of recombination sites are particularly suitable for
engineering as cellular protein production systems.

In one embodiment, CPP *ACes* involve a two-component system:
25 the platform chromosome containing multiple engineering sites and the
donor target vector containing a platform-specific recombination site with
designed expression cassettes (see Figure 9).

The platform *ACes* can be produced from any artificial
chromosome, particularly the amplification-based artificial chromosomes.

For exemplification, they are produced from rodent artificial chromosomes produced from acrocentric chromosomes using the technology of U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183, in which nucleic acid is targeted to the pericentric heterochromatic, and, particularly into rDNA to initiate the replication event(s). The ACes can be produced directly in the chosen cellular protein production cell lines, such as, but not limited to, CHO cells, hybridomas, plant cells, plant tissues, plant protoplasts, stem cells and plant calli.

10 a. **Platform Construction**

 In the exemplary embodiment, the initial *de novo* platform construction requires co-transfecting with excess targeting DNA, such as, rDNA or lambda DNA without an *attP* region, and an engineered selectable marker. The engineered selectable marker should contain promoter, generally a constitutive promoter, such as human, viral, i.e., adenovirus or SV40 promoter, including the human ferritin heavy chain promoter (SEQ ID NO:128), SV40 and EF1 α promoters, to control expression of a marker gene that provides a selective growth advantage to the cell. An example of such a marker gene is the *E. coli hisD* gene (encoding histidinol dehydrogenase) which is homologous and analogous to the *S. typhimurium hisD* a dominant marker selection system for mammalian cells previously described (see, Hartman *et al.* (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85:8047-8051). Since histidine is an essential amino acid in mammals and a nutritional requirement in cell culture, the *E. coli hisD* gene can be used to select for histidine prototrophy in defined media. Furthermore more stringent selection can be placed on the cells by including histidinol in the medium. Histidinol is itself permeable and toxic to cells. The *hisD* provides a means of detoxification.

Placed between the promoter and the marker gene is the bacteriophage lambda *attP* site to use the bacteriophage lambda integrase dependent site-specific recombination system (described herein). The insertion of an *attP* site downstream of a promoter element provide
5 forward selection of site-specific recombination events onto the platform *ACes*.

b. Donor Target Vector Construction

A second component of the CPP platform *ACes* system involves the construction of donor target vectors containing a gene product(s) of
10 interest for the CPP platform *ACes*. Individual donor target vectors can be designed for each gene product to be expressed thus enabling maximum usage of a *de novo* constructed platform *ACes*, so that one or a few CPP platform *ACes* will be required for many gene targets.

A key feature of the donor vector target is the *promoterless* marker
15 gene containing an upstream *attB* site (marker 2 on figure 9). Normally the marker would not be expressed unless it is placed downstream of a promoter sequence. As discussed above, using the lambda integrase technology (λ INT_{E174R} on Figure 8 and Figure 9), site-specific recombination between the *attB* site on the vector and the promoter-*attP*
20 site on the CPP platform *ACes* result in the expression of the donor target vector marker providing positive selection for the site-specific event. Site-specific recombination events on the CPP *ACes* versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR.
25 In addition, since the lambda integrase reaction is unidirectional, i.e. excision reaction is not possible, a number of unique targets can be loaded onto the CPP platform *ACes* limited only by the number of markers available.

Additional features of the donor target vector include gene target expression cassettes flanked by either chromatin insulator regions, matrix attachment regions (MAR) or scaffold attachment regions (SAR). The use of these regions will provide a more "open" chromatin environment for gene expression and help alleviate silencing. An example of such a cassette for expressing a monoclonal antibody is described. For this purpose, a strong constitutive promoter, e.g. chicken β -actin or RNA PolII, is used to drive the expression of the heavy and light chain open reading frames. The heavy and light chain sequences flank a nonattenuated human IRES (IRES_H; from the 5'UTR of NRF1 gene; see Oumard et al., 2000, *Mol. and Cell Biol.*, 20(8):2755-2759) element thereby coordinating transcription of both heavy and light chain sequence. Distal to the light chain open reading frame resides an additional viral encoded IRES (IRES_V modified ECMV internal ribosomal entry site (IRES)) element attenuating the expression of the fluorescent marker gene hrGFP from *Renilla* (Stratagene). By linking the hrGFP with an attenuated IRES, the heavy and light chains along with the hrGFP are monocistronic. Thus, the identification of hrGFP fluorescing cells will provide a means to detect protein producing cells. In addition, high producing cell lines can be identified and isolated by FACS thereby decreasing the time frame in finding high expressers. Functional monoclonal antibody will be confirmed by ELISA.

c. Additional components in cellular protein production platform ACes (CPP ACes)

In addition to the aforementioned CPP ACes system, other genetic factors can be included to enhance the yield and quality of the expressed protein. Again to provide maximum flexibility, these additional factors can be inserted onto the CPP platform ACes by λ INTE174R dependent site-specific recombination. Other factors that could be used with a CPP

Platform *ACes* include for example, adenovirus E1a transactivation system which upregulates both cellular and viral promoters (see, e.g., Svensson and Akusjarvi (1984) EMBO 3:789-794; and US patents 5,866,359; 4,775,630 and 4,920,211).

5 **d. Targets for CHO-*ACes* engineering to enhance cell growth, such as CHO cell growth and protein production/ quality**

 If adding these additional factors onto the CPP *ACes* is not prudent or desired, the host cell, CHO cells, can be engineered to express these
10 factors (see, below, targets for CHO-*ACes* engineering to enhance CHO cell growth and protein production/quality). Additional factors to consider including are addition of insulin or IGF-1 to sustain viability; human sialyltransferases or related factors to produce more human-like
15 glycoproteins; expression of factors to decrease ammonium accumulation during cell growth; expression of factors to inhibit apoptosis; expression of factors to improve protein secretion and protein folding; and expression of factors to permit serum-free transfection and selection.

1) Addition of insulin or IGF-1 to sustain viability

20 Stimulatory factors and/or their receptors are expressed to set up an autocrine loop, to improve cell growth, such as CHO cell growth. Two exemplary candidates are insulin and IGF-1 (see, Biotechnol Prog 2000 Sep;16(5):693-7). Insulin is the most commonly used growth factor for sustaining cell growth and viability in serum-free Chinese hamster ovary
25 (CHO) cell cultures. Insulin and IGF-1 analog (LongR(3) serve as growth and viability factors for CHO cells.

 CHO cells were modified to produce higher levels of essential nutrients and factors. A serum-free (SF) medium for dihydrofolate reductase-deficient Chinese hamster ovary cells (DG44 cells) was
30 prepared. Chinese hamster ovary cells (DG44 cells), which are normally

maintained in 10% serum medium, were gradually weaned to 0.5% serum medium to increase the probability of successful growth in SF medium (see, Kim *et al.* (199) *In Vitro Cell Dev Biol Anim* 35(4):178-82). A SF medium (SF-DG44) was formulated by supplementing the basal medium with these components; basal medium was prepared by supplementing Dulbecco's modified Eagle's medium and Ham's nutrient mixture F12 with hypoxanthine (10 mg/l) and thymidine (10 mg/l). Development of a SF medium for DG44 cells was facilitated using a Plackett-Burman design technique and weaning of cells.

10

2) Human sialyltransferases or related factors to produce more human-like glycoproteins

CHO cells have been modified by increasing their ability to process protein via addition of complex carbohydrates. This has been achieved by overexpression of relevant processing enzymes, or in some cases, reducing expression of relevant enzymes (see, Bragonzi *et al.* (2000) *Biochim Biophys Acta* 1474(3):273-282; see, also Weikert *et al.* (1999) *Nature biotech.* 17:1116-11121; Ferrari J *et al.* (1998) *Biotechnol Bioeng* 60(5):589-95). A CHO cell line expressing alpha2,6-sialyltransferase was developed for the production of human-like sialylated recombinant glycoproteins. The sialylation defect of CHO cells can be corrected by transfecting the alpha2,6-sialyltransferase (alpha2,6-ST) cDNA into the cells. Glycoproteins produced by such CHO cells display alpha2,6-and alpha2,3-linked terminal sialic acid residues, similar to human glycoproteins.

As another example for improving the production of human-like sialylated recombinant glycoproteins, a CHO cell line has been developed that constitutively expresses sialidase antisense RNA (see, Ferrari J *et al.* (1998) *Biotechnol Bioeng* 60(5):589-95). Several antisense expression

30

vectors were prepared using different regions of the sialidase gene. Co-transfection of the antisense constructs with a vector conferring puromycin resistance gave rise to over 40 puromycin resistant clones that were screened for sialidase activity. A 5' 474 bp coding segment of the sialidase cDNA, in the inverted orientation in an SV 40-based expression vector, gave maximal reduction of the sialidase activity to about 40% wild-type values.

Oligosaccharide biosynthesis pathways in mammalian cells have been engineered for generation of recombinant glycoproteins (see, *e.g.*, Sburlati (1998) *Biotechnol Prog* 14(2):189-92), which describes a Chinese hamster ovary (CHO) cell line capable of producing bisected oligosaccharides on glycoproteins. This cell line was created by overexpression of a recombinant N-acetylglucosaminyltransferase III (GnT-III) (see, also, Prati *et al.* (1998) *Biotechnol Bioeng* 59(4):445-50, which describes antisense strategies for glycosylation engineering of CHO cells).

3) Expression of factors to decrease ammonium accumulation during cell growth

Excess ammonium, which is a by-product of CHO cell metabolism can have detrimental effects on cell growth and protein quality (see, Yang *et al.* (2000) *Biotechnol Bioeng* 68(4):370-80). To solve this problem ammonium levels were modified by overexpressing carbamoyl phosphate synthetase I and ornithine transcarbamoylase or glutamine synthetase in CHO cells. Such modification resulted in reduced ammonium levels observed and an increase in the growth rate (see Kim *et al.* (2000) *J Biotechnol* 81(2-3):129-40; and Enosawa *et al.* (1997) *Cell Transplant* 6(5):537-40).

4) Expression of factors to improve protein secretion and protein folding

Overexpression of relevant enzymes can be engineered into the ACes to improve protein secretion and folding.

5) Expression of factors to permit serum-free transfection and selection

5 It is advantageous to have the ability to convert CHO cells in suspension growing in serum free medium to adherence with out having to resort to serum addition. Laminin or fibronectin addition is sufficient to make cells adherent (see, *e.g.*, Zaworski *et al.* (1993) *Biotechniques* 15(5):863-6) so that expressing either of these genes in CHO cells under
10 an inducible promoter should allow for reversible shift to adherence without requiring serum addition.

2. Platform ACes and Gene Therapy

The platform ACes provided herein are contemplated for use in mammalian gene therapy, particularly human gene therapy. Human ACes
15 can be derived from human acrocentric chromosomes from human host cells, in which the amplified sequences are heterochromatic and/or human rDNA. Different platform ACes applicable for different tissue cell types are provided. The ACes for gene therapy can contain a single copy of a therapeutic gene inserted into a defined location on platform ACes.
20 Therapeutic genes include genomic clones, cDNA, hybrid genes and other combinations of sequences. Preferred selectable markers are those from the mammalian host, such as human derived factors so that they are non-immunogenic, non-toxic and allow for efficient selection, such as by FACS and/or drug resistance.

25 Platform ACes, useful for gene therapy and other applications, as noted herein, can be generated by megareplicator dependent amplification, such as by the methods in U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183. In one embodiment, human ACes are produced using

human rDNA constructs that target rDNA arrays on human acrocentric chromosomes and induce the megareplicator in human cells, particularly in primary cell lines (with sufficient number of doublings to form the ACes) or stem cells (such as hematopoietic stem cells, mesenchymal stem cells, adult stem cells or embryonic stem cells) to avoid the introduction of potentially harmful rearranged DNA sequences present in many transformed cell lines. Megareplicator induced ACes formation can result in multiple copies of targeting DNA/selectable markers in each amplification block on both chromosomal arms of the platform ACes.

10 In view of the considerations regarding immunogenicity and toxicity, the production of human platform ACes for gene therapy applications employs a two component system analogous to the platform ACes designed for cellular protein production (CPP platform ACes). The system includes a platform chromosome of entirely human DNA origin
15 containing multiple engineering sites and a gene target vector carrying the therapeutic gene of interest.

a. Platform Construction

The initial *de novo* construction of the platform chromosome employs the co-transfection of excess targeting DNA and a selectable
20 marker. In one embodiment, the DNA is targeted to the rDNA arrays on the human acrocentric chromosomes (chromosomes 13, 14, 15, 21 and 22). For example, two large human rDNA containing PAC clones 18714 and 18720 and the human PAC clone 558F8 are used for targeting (Genome Research (ML) now Incyte, BACPAC Resources, 747 52nd
25 Street, Oakland CA). The mouse rDNA clone pFK161 (SEQ ID NO: 118), which was used to make the human SATAC from the 94-3 hamster/human hybrid cell line (see, *e.g.*, published International PCT application No. WO 97/40183 and Csonka, *et al*, *Journal of Cell Science*

113:3207-32161 and Example 1 for a description of pFK161) can also be used.

For animal applications, selectable markers should be non-immunogenic in the animal, such as a human, and include, but are not limited to: human nerve growth factor receptor (detected with a MAb, such as described in US patent 6,365,373); truncated human growth factor receptor (detected with MAb), mutant human dihydrofolate reductase (DHFR; fluorescent MTX substrate available); secreted alkaline phosphatase (SEAP; fluorescent substrate available); human thymidylate synthase (TS; confers resistance to anti-cancer agent fluorodeoxyuridine); human glutathione S-transferase alpha (GSTA1; conjugates glutathione to the stem cell selective alkylator busulfan; chemoprotective selectable marker in CD34+ cells); CD24 cell surface antigen in hematopoietic stem cells; human CAD gene to confer resistance to N-phosphonacetyl-L-aspartate (PALA); human multi-drug resistance-1 (MDR-1; P-glycoprotein surface protein selectable by increased drug resistance or enriched by FACS); human CD25 (IL-2 α ; detectable by Mab-FITC); Methylguanine-DNA methyltransferase (MGMT; selectable by carmustine); and Cytidine deaminase (CD; selectable by Ara-C).

Since megareplicator induced amplification generates multiple copies of the selectable marker, a second consideration for the selection of the human marker is the resulting dose of the expressed marker after ACes formation. High level of expression of certain markers may be detrimental to the cell and/or result in autoimmunity. One method to decrease the dose of the marker protein is by shortening its half-life, such as via the fusion of the well-conserved human ubiquitin tag (a 76 amino acid sequence) thus leading to increased turnover of the selectable marker. This has been used successfully for a number of reporter

systems including DHFR (see, *e.g.*, Stack *et al.* (2000) *Nature Biotechnology* 18:1298-1302 and references cited therein).

Using the ubiquitin tagged protein, a human selectable marker system analogous to the CPP ACes described herein is constructed.

- 5 Briefly, a tagged selectable marker, such as for example one of those described herein, is cloned downstream of an *attP* site and expressed from a human promoter. Exemplary promoters contemplated for use herein include, but are not limited to, the human ferritin heavy chain promoter (SEQ ID NO:128); RNA Poll; EF1 α ; TR; glyceraldehyde-3-
10 phosphate dehydrogenase core promoter (GAP); a GAP core promoter including a proximal insulin inducible element and the intervening GAP sequence; phosphofructokinase promoter; and phosphoglycerate kinase promoter. Also contemplated herein is an aldolase A promoter H1 & H2 (representing closely spaced transcriptional start sites) along with the
15 proximal H enhancer. There are 4 promoters (*e.g.*, transcriptional start sites) for this gene, each having different regulatory and tissue activity. The H (most proximal 2) promoters are ubiquitously expressed off the H enhancer. This resulting marker can then be co-transfected along with excess human rDNA targeting sequence into the host cells. An important
20 criteria for the selection of the

- recipient cells is sufficient number of cell doublings for the formation and detection of ACes. Accordingly, the co-transfections should be attempted in human primary cells that can be cultured for long periods of
25 time, such as for example, stem cells (*e.g.*, hematopoietic, mesenchymal, adult or embryonic stem cells), or the like. Additional cell types, include, but are not limited to: single gene transfected cells exhibiting increased life-span; over-expressing c-myc cells, *e.g.* MSU1.1 (Morgan *et al.*, 1991, *Exp. Cell Res.*, Nov;197(1):125-136); over-expressing telomerase lines,

such as TERT cells; SV40 large T-antigen transfected lines; tumor cell lines, such as HT1080; and hybrid human cell lines, such as the 94-3 hamster/human hybrid cell line.

b. Gene Target Vector

5 The second component of the GT platform *ACes* (GT *ACes*) system involves the use of engineered target vectors carrying the therapeutic gene of interest. These are introduced onto the GT platform *ACes* via site-specific recombination. As with the CPP *ACes*, the use of engineered target vectors maximizes the use of the *de novo* generated GT platform
10 *ACes* for most gene targets. Furthermore, using lambda integrase technology, GT platform *ACes* containing multiple *attP* sites permits the opportunity to incorporate multiple therapeutic targets onto a single platform. This could be of value in cases where a defined therapy requires multiple gene targets, a single therapeutic target requires an
15 additional gene regulatory factor or a GT *ACes* requires a "kill" switch.

Similar to the CPP *ACes*, a feature of the gene target vector is the *promoterless* marker gene containing an upstream *attB* site (marker 2 on Figure 9). Normally, the marker (in this case, a cell surface antigen that can be sorted by FACS would be ideal) would not be expressed unless it
20 is placed downstream of a promoter sequence. Using the lambda integrase technology (λ INT_{E174R} on figure 9), site-specific recombination between the *attB* site on the vector and the promoter- *attP* site on the GT platform *ACes* results in the expression of marker#2 on the gene target vector, i.e. positive selection for the site-specific event. Site-specific
25 recombination events on the GT *ACes* versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR.

For expression of the therapeutic gene, human specific promoters, such as a ferritin heavy chain promoter (SEQ ID NO:128); EF1 α or RNA

Poll, are used. These promoters are for high level expression of a cDNA encoded therapeutic protein. In addition to expressing cDNA (or even hybrid cDNA/artificial intron constructs), the GT platform *ACes* are used for engineering and expressing large genomic fragments carrying
5 therapeutic genes of interest expressed from native promoter sequences. This is of importance in situations where the therapy requires precise cell specific expression or in instances where expression is best achieved from genomic clones versus cDNA.

10 **3. Selectable markers for use, for example, in Gene Therapy (GT)**

 The following are selectable markers that can be incorporated into human *ACes* and used for selection.

15 **Dual Resistance to 4-Hydroperoxycyclophosphamide and Methotrexate by Retroviral Transfer of the Human Aldehyde Dehydrogenase Class 1 Gene and a Mutated Dihydrofolate Reductase Gene**

 The genetic transfer of drug resistance to hematopoietic cells is one approach to overcoming myelosuppression caused by high-dose chemotherapy. Because cyclophosphamide (CTX) and methotrexate
20 (MTX) are commonly used non-cross-resistant drugs, generation of dual drug resistance in hematopoietic cells that allows dose intensification may increase anti-tumor effects and circumvent the emergence of drug-resistant tumors, a retroviral vector containing a human cytosolic ALDH-1-encoding DNA clone and a human doubly mutated DHFR-encoding
25 clone (Phe22/Ser31; termed F/S in the description of constructs) to generate increased resistance to CTX and MTX were constructed (Takebe *et al.* (2001) *Mol Ther* 3(1):88-96). This construct may be useful for protecting patients from high-dose CTX- and MTX-induced myelosuppression. *ACes* can be similarly constructed.

Multiple mechanisms of N-phosphonacetyl-L-aspartate resistance in human cell lines: carbamyl-P synthetase/aspartate transcarbamylase/dihydro-orotase gene amplification is frequent only when chromosome 2 is rearranged

5

Rodent cells resistant to N-phosphonacetyl-L-aspartate (PALA) invariably contain amplified carbamyl-P synthetase/aspartate transcarbamylase/dihydro-orotase (CAD) genes, usually in widely spaced tandem arrays present as extensions of the same chromosome arm that carries a single copy of CAD in normal cells (Smith *et al.* (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:1816-21). In contrast, amplification of CAD is very infrequent in several human tumor cell lines. Cell lines with minimal chromosomal rearrangement and with unrearranged copies of chromosome 2 rarely develop intrachromosomal amplifications of CAD. These cells frequently become resistant to PALA through a mechanism that increases the aspartate transcarbamylase activity with no increase in CAD copy number, or they obtain one extra copy of CAD by forming an isochromosome 2p or by retaining an extra copy of chromosome 2. In cells with multiple chromosomal aberrations and rearranged copies of chromosome 2, amplification of CAD as tandem arrays from rearranged chromosomes is the most frequent mechanism of PALA resistance. All of these different mechanisms of PALA resistance are blocked in normal human fibroblasts. Thus, ACes with multiple copies of the CAD gene would provide PALA resistance.

25

Retroviral coexpression of thymidylate synthase and dihydrofolate reductase confers fluoropyrimidine and antifolate resistance

Retroviral gene transfer of dominant selectable markers into hematopoietic cells can be used to select genetically modified cells in vivo or to attenuate the toxic effects of chemotherapeutic agents. Fantz *et al.* ((1998) *Biochem Biophys Res Comm* 243(1):6-12) have shown that

30

retroviral gene transfer of thymidylate synthase (TS) confers resistance to TS directed anticancer agents and that co-expression of TS and dihydrofolate reductase (DHFR) confers resistance to TS and DHFR cytotoxic agents. Retroviral vectors encoding *Escherichia coli* TS, human
5 TS, and the Tyr-to-His at residue 33 variant of human TS (Y33HhTS) were constructed and fibroblasts transfected with these vectors conferred comparable resistance to the TS-directed agent fluorodeoxyuridine (FdUrd, approximately 4-fold). Retroviral vectors that encode dual
10 expression of Y33HhTS and the human L22Y DHFR (L22YhDHFR) variants conferred resistance to FdUrd (3- to 5-fold) and trimetrexate (30- to 140-fold). A L22YhDHFR-Y33HhTS chimeric retroviral vector was also constructed and transduced cells were resistant to FdUrd (3-fold), AG337 (3-fold), trimetrexate (100-fold) and methotrexate (5-fold). These results show that recombinant retroviruses can be used to transfer the cDNA
15 that encodes TS and DHFR and dual expression in transduced cells is sufficiently high to confer resistance to TS and DHFR directed anticancer agents. ACes can be similarly constructed.

Human CD34 + cells do not express glutathione S-transferases alpha

20 The expression of glutathione S-transferases alpha (GST alpha) in human hematopoietic CD34 + cells and bone marrow was studied using RT-PCR and immunoblotting (Czerwinski M, Kiem *et al.* (1997) *Gene Ther* 4(3):268-70). The GSTA1 protein conjugates glutathione to the stem cell selective alkylator busulfan. This reaction is the major pathway of
25 elimination of the compound from the human body. Human hematopoietic CD34 + cells and bone marrow do not express GSTA1 message, which was present at a high level in liver, an organ relatively resistant to busulfan toxicity in comparison to bone marrow. Similarly, baboon CD34 + cells and dog bone marrow do not express GSTA1. Thus, human

GSTA1 is a chemoprotective selectable marker in human stem cell gene therapy and could be employed in ACes construction.

Selection of retrovirally transduced hematopoietic cells using CD24 as a marker of gene transfer

5 Pawliuk *et al.* ((1994) *Blood* 84(9):2868-2877) have investigated the use of a cell surface antigen as a dominant selectable marker to facilitate the detection and selection of retrovirally infected target cells. The small coding region of the human cell surface antigen CD24 (approximately 240 bp) was introduced into a myeloproliferative sarcoma virus (MPSV)-based retroviral vector, which was then used to infect day 4
10 5-fluorouracil (5-FU)-treated murine bone marrow cells. Within 48 hours of termination of the infection procedure CD24-expressing cells were selected by fluorescent-activated cell sorting (FACS) with an antibody directed against the CD24 antigen. Functional analysis of these cells
15 showed that they included not only in vitro clonogenic progenitors and day 12 colony-forming unit-spleen but also cells capable of competitive long-term hematopoietic repopulation. Double-antibody labeling studies performed on recipients of retrovirally transduced marrow cells showed that some granulocytes, macrophages, erythrocytes, and, to a lesser
20 extent, B and T lymphocytes still expressed the transduced CD24 gene at high levels 4 months later. No gross abnormalities in hematopoiesis were detected in mice repopulated with CD24-expressing cells. These results show that the use of the CD24 cell surface antigen as a retrovirally encoded marker permits rapid, efficient, and nontoxic selection *in vitro* of
25 infected primary cells, facilitates tracking and phenotyping of their progeny, and provides a tool to identify elements that regulate the expression of transduced genes in the most primitive hematopoietic cells. ACes could be similarly constructed.

DeltahGHR, a biosafe cell surface-labeling molecule for analysis and selection of genetically transduced human cells

A selectable marker for retroviral transduction and selection of human and murine cells is known (see, Garcia-Ortiz *et al.* (2000) *Hum Gene Ther* 11(2):333-46). The molecule expressed on the cell surface of the transduced population is a truncated version of human growth hormone receptor (deltahGHR), capable of ligand (hGH) binding, but devoid of the domains involved in signal triggering. The engineered molecule is stably expressed in the target cells as an inert protein unable to trigger proliferation or to rescue the cells from apoptosis after ligand binding. This new marker, has a wide application spectrum, since hGHR in the human adult is highly expressed only in liver cells, and lower levels have been reported in certain lymphocyte cell populations. The deltagHHR label has high biosafety potential, as it belongs to a well-characterized hormonal system that is nonessential in adults, and there is extensive clinical experience with hGH administration in humans. The differential binding properties of several monoclonal antibodies (MAbs) are used in a cell rescue method in which the antibody used to select deltagHHR-transduced cells is eluted by competition with hGH or, alternatively biotinylated hGH is used to capture tagged cells. In the latter system, the final purified population is recovered free of attached antibodies in hGH (a substance approved for human use)-containing medium. Such a system could be used to identify ACes containing cells.

4. **Transgenic models for evaluation of genes and discovery of new traits in plants**

Of interest is the use of plants and plant cells containing artificial chromosomes for the evaluation of new genetic combinations and discovery of new traits. Artificial chromosomes, by virtue of the fact that they can contain significant amounts of DNA can also therefore encode

numerous genes and accordingly a multiplicity of traits. It is contemplated here that artificial chromosomes, when formed from one plant species, can be evaluated in a second plant species. The resultant phenotypic changes observed, for example, can indicate the nature of the genes contained within the DNA contained within the artificial chromosome, and hence permit the identification of novel genetic activities. Artificial chromosomes containing euchromatic DNA or partially containing euchromatic DNA can serve as a valuable source of new traits when transferred to an alien plant cell environment. For example, it is contemplated that artificial chromosomes derived from dicot plant species can be introduced into monocot plant species by transferring a dicot artificial chromosome. The dicot artificial chromosome possessing a region of euchromatic DNA containing expressed genes.

The artificial chromosomes can be designed to allow the artificial chromosome to recombine with the naturally occurring plant DNA in such a fashion that a large region of naturally occurring plant DNA becomes incorporated into the artificial chromosome. This allows the artificial chromosome to contain new genetic activities and hence carry novel traits. For example, an artificial chromosome can be introduced into a wild relative of a crop plant under conditions whereby a portion of the DNA present in the chromosomes of the wild relative is transferred to the artificial chromosome. After isolation of the artificial chromosome, this naturally occurring region of DNA from the wild relative, now located on the artificial chromosome can be introduced into the domesticated crop species and the genes encoded within the transferred DNA expressed and evaluated for utility. New traits and gene systems can be discovered in this fashion. The artificial chromosome can be modified to contain sequences that promote homologous recombination within plant cells, or

be modified to contain a genetic system that functions as a site-specific recombination system.

Artificial chromosomes modified to recombine with plant DNA offer many advantages for the discovery and evaluation of traits in different
5 plant species. When the artificial chromosome containing DNA from one plant species is introduced into a new plant species, new traits and genes can be introduced. This use of an artificial chromosome allows for the ability to overcome the sexual barrier that prevents transfer of genes from one plant species to another species. Using artificial chromosomes in this
10 fashion allows for many potentially valuable traits to be identified including traits that are typically found in wild species. Other valuable applications for artificial chromosomes include the ability to transfer large regions of DNA from one plant species to another, such as DNA encoding potentially valuable traits such as altered oil, carbohydrate or protein
15 composition, multiple genes encoding enzymes capable of producing valuable plant secondary metabolites, genetic systems encoding valuable agronomic traits such as disease and insect resistance, genes encoding functions that allow association with soil bacterium such as growth promoting bacteria or nitrogen fixing bacteria, or genes encoding traits
20 that confer freezing, drought or other stress tolerances. In this fashion, artificial chromosomes can be used to discover regions of plant DNA that encode valuable traits.

The artificial chromosome can also be designed to allow the transfer and subsequent incorporation of these valuable traits now located
25 on the artificial chromosome into the natural chromosomes of a plant species. In this fashion the artificial chromosomes can be used to transfer large regions of DNA encoding traits normally found in one plant species into another plant species. In this fashion, it is possible to derive a plant cell that no longer needs to carry an artificial chromosome to

posses the novel trait. Thus, the artificial chromosome would serve as the transfer mechanism to permit the formation of plants with greater degree of genetic diversity.

5 The design of an artificial chromosome to accomplish the afore-
mentioned purposes can include within the artificial chromosome the
presence of specific DNA sequences capable of acting as sites for
homologous recombination to take place. For example, the DNA
sequence of *Arabidopsis* is now known. To construct an artificial
chromosome capable of recombining with a specific region of *Arabidopsis*
10 DNA, a sequence of *Arabidopsis* DNA, normally located near a
chromosomal location encoding genes of potential interest can be
introduced into an artificial chromosome by methods provided herein. It
may be desirable to include a second region of DNA within the artificial
chromosome that provides a second flanking sequence to the region
15 encoding genes of potential interest, to promote a double recombination
event which would ensure transfer of the entire chromosomal region,
encoding genes of potential interest, to the artificial chromosome. The
modified artificial chromosome, containing the DNA sequences capable of
homologous recombination region, can then be introduced into
20 *Arabidopsis* cells and the homologous recombination event selected.

It is convenient to include a marker gene to allow for the selection
of a homologous recombination event. The marker gene is preferably
inactive unless activated by an appropriate homologous recombination
event. For example, US 5,272,071, describes a method where an
25 inactive plant gene is activated by a recombination event such that
desired homologous recombination events can be easily scored. Similarly,
US 5,501,967 describes a method for the selection of homologous
recombination events by activation of a silent selection gene first
introduced into the plant DNA, the gene being activated by an appropriate

homologous recombination event. Both of these methods can be applied to enable a selective process to be included to select for recombination between an artificial chromosome and a plant chromosome. Once the homologous recombination event is detected, the artificial chromosome, once selected, is isolated and introduced into a recipient cell, for example, tobacco, corn, wheat or rice, and the expression of the newly introduced DNA sequences evaluated.

Phenotypic changes in the recipient plant cells containing the artificial chromosome, or in regenerated plants containing the artificial chromosome, allows for the evaluation of the nature of the traits encoded by the *Arabidopsis* DNA, under conditions naturally found in plant cells, including the naturally occurring arrangement of DNA sequences responsible for the developmental control of the traits in the normal chromosomal environment.

Traits such as durable fungal or bacterial disease resistance, new oil and carbohydrate compositions, valuable secondary metabolites such as phytosterols, flavonoids, efficient nitrogen fixation or mineral utilization, resistance to extremes of drought, heat or cold are all found within different populations of plant species and are often governed by multiple genes. The use of single gene transformation technologies does not permit the evaluation of the multiplicity of genes controlling many valuable traits. Thus, incorporation of these genes into artificial chromosomes allows the rapid evaluation of the utility of these genetic combinations in heterologous plant species.

The large scale order and structure of the artificial chromosome provides a number of unique advantages in screening for new utilities or novel phenotypes within heterologous plant species. The size of new DNA that can be carried by an artificial chromosome can be millions of base pairs of DNA, representing potentially numerous genes that may

have novel utility in a heterologous plant cell. The artificial chromosome is a "natural" environment for gene expression, the problems of variable gene expression and silencing seen for genes transferred by random insertion into a genome should not be observed. Similarly, there is no
5 need to engineer the genes for expression, and the genes inserted would not need to be recombinant genes. Thus, one expects the expression from the transferred genes to be temporal and spatial, as observed in the species from where the genes were initially isolated. A valuable feature for these utilities is the ability to isolate the artificial chromosomes and to
10 further isolate, manipulate and introduce into other cells artificial chromosomes carrying unique genetic compositions.

Thus, the use of artificial chromosomes and homologous recombination in plant cells can be used to isolate and identify many valuable crop traits.

15 In addition to the use of artificial chromosomes for the isolation and testing of large regions of naturally occurring DNA, methods for the use of artificial chromosomes and cloned DNA are also contemplated. Similar to that described above, artificial chromosomes can be used to carry large regions of cloned DNA, including that derived from other plant species.

20 The ability to incorporate novel DNA elements into an artificial chromosome as it is being formed allows for the development of artificial chromosomes specifically engineered as a platform for testing of new genetic combinations, or "genomic" discoveries for model species such as *Arabidopsis*. It is known that specific "recombinase" systems can be
25 used in plant cells to excise or re-arrange genes. These same systems can be used to derive new gene combinations contained on an artificial chromosome.

The artificial chromosomes can be engineered as platforms to accept large regions of cloned DNA, such as that contained in Bacterial

Artificial Chromosomes (BACs) or Yeast Artificial Chromosomes (YACs). It is further contemplated, that as a result of the typical structure of artificial chromosomes containing tandemly repeated DNA blocks, that sequences other than cloned DNA sequence can be introduced by recombination processes. In particular recombination within a predefined region of the tandemly repeated DNA within the artificial chromosome provides a mechanism to "stack" numerous regions of cloned DNA, including large regions of DNA contained within BACs or YACs clones. Thus, multiple combinations of genes can be introduced onto artificial chromosomes and these combinations tested for functionality. In particular, it is contemplated that multiple YACs or BACs can be stacked onto an artificial chromosomes, the BACs or YACs containing multiple genes of complex pathways or multiple genetic pathways. The BACs or YACs are typically selected based on genetic information available within the public domain, for example from the Arabidopsis Information Management System (<http://aims.cps.msu.edu/aims/index.html>) or the information related to the plant DNA sequences available from the Institute for Genomic Research (<http://www.tigr.org>) and other sites known to those skilled in the art. Alternatively, clones can be chosen at random and evaluated for functionality. It is contemplated that combinations providing a desired phenotype can be identified by isolation of the artificial chromosome containing the combination and analyzing the nature of the inserted cloned DNA.

In this regard, it is contemplated that the use of site-specific recombination sequences can have considerable utility in developing artificial chromosomes containing DNA sequences recognized by recombinase enzymes and capable of accepting DNA sequences containing same. The use of site-specific recombination as a means to target an introduced DNA to a specific locus has been demonstrated in

the art and such methods can be employed. The recombinase systems can also be used to transfer the cloned DNA regions contained within the artificial chromosome to the naturally occurring plant or mammalian chromosomes.

5 As noted herein, many site-specific recombinases are known and can be identified (Kilby *et al.* (1993) *Trends in Genetics* 9:413-418). The three recombinase systems that have been extensively employed include: an activity identified as R encoded by the pSR1 plasmid of *Zygosaccharomyces rouxii*, FLP encoded for the 2 μ m circular plasmid from
10 *Saccharomyces cerevisiae* and *Cre-lox* from the phage P1.

 The integration function of site-specific recombinases is contemplated as a means to assist in the derivation of genetic combinations on artificial chromosomes. In order to accomplish this, it is contemplated that a first step of introducing site-specific recombinase
15 sites into the genome of a plant cell in an essentially random manner is conducted, such that the plant cell has one or more site-specific recombinase recognition sequences on one or more of the plant chromosomes. An artificial chromosome is then introduced into the plant cell, the artificial chromosome engineered to contain a recombinase
20 recognition site (e.g., integration site) capable of being recognized by a site-specific recombinase. Optionally, a gene encoding a recombinase enzyme is also included, preferably under the control of an inducible promoter. Expression of the site-specific recombinase enzyme in the plant cell, either by induction of a inducible recombinase gene, or
25 transient expression of a recombinase sequence, causes a site-specific recombination event to take place, leading to the insertion of a region of the plant chromosomal DNA (containing the recombinase recognition site) into the recombinase recognition site of the artificial chromosome, and forming an artificial chromosome containing plant chromosomal DNA.

The artificial chromosome can be isolated and introduced into a heterologous host, preferably a plant host, and expression of the newly introduced plant chromosomal DNA can be monitored and evaluated for desirable phenotypic changes. Accordingly, carrying out this

5 recombination with a population of plant cells wherein the chromosomally located recombinase recognition site is randomly scattered throughout the chromosomes of the plant, can lead to the formation of a population of artificial chromosomes, each with a different region of plant chromosomal DNA, and each potentially representing a novel genetic combination.

10 This method requires the precise site-specific insertion of chromosomal DNA into the artificial chromosome. This precision has been demonstrated in the art. For example, Fukushige and Sauer ((1992) Proc. Natl. Acad. Sci. USA, 89:7905-7909) demonstrated that the *Cre-lox* homologous recombination system could be successfully employed to
15 introduce DNA into a predefined locus in a chromosome of mammalian cells. In this demonstration a promoter-less antibiotic resistance gene modified to include a *lox* sequence at the 5' end of the coding region was introduced into CHO cells. Cells were re-transformed by electroporation with a plasmid that contained a promoter with a *lox* sequence and a
20 transiently expressed *Cre* recombinase gene. Under the conditions employed, the expression of the *Cre* enzyme catalyzed the homologous recombination between the *lox* site in the chromosomally located promoter-less antibiotic resistance gene, and the *lox* site in the introduced promoter sequence, leading to the formation of a functional antibiotic
25 resistance gene. The authors demonstrated efficient and correct targeting of the introduced sequence, 54 of 56 lines analyzed corresponded to the predicted single copy insertion of the DNA due to *Cre* catalyzed site-specific homologous recombination between the *lox* sequences.

Accordingly a *lox* sequence may be first added to a genome of a plant species capable of being transformed and regenerated to a whole plant to serve as a recombinase target DNA sequence for recombination with an artificial chromosome. The *lox* sequence may be optimally
5 modified to further contain a selectable marker which is inactive but can be activated by insertion of the *lox* recombinase recognition sequence into the artificial chromosome.

A promoterless marker gene or selectable marker gene linked to the recombinase recognition sequence, which is first inserted into the
10 chromosomes of a plant cell can be used to engineer a platform chromosome. A promoter is linked to a recombinase recognition site, in an orientation that allows the promoter to control the expression of the marker or selectable marker gene upon recombination within the artificial chromosome. Upon a site-specific recombination event between a
15 recombinase recognition site in a plant chromosome and the recombinase recognition site within the introduced artificial chromosome, a cell is derived with a recombined artificial chromosome, the artificial chromosome containing an active marker or selectable marker activity that permits the identification and or selection of the cell.

20 The artificial chromosomes can be transferred to other plant or animal species and the functionality of the new combinations tested. The ability to conduct such an inter-chromosomal transfer of sequences has been demonstrated in the art. For example, the use of the *Cre-lox* recombinase system to cause a chromosome recombination event
25 between two chromatids of different chromosomes has been shown.

Any number of recombination systems may be employed as described herein, such as, but not limited to, bacterially derived systems such as the att/int system of phage lambda, and the Gin/gix system.

More than one recombination system may be employed, including, for example, one recombinase system for the introduction of DNA into an artificial chromosome, and a second recombinase system for the subsequent transfer of the newly introduced DNA contained within an
5 artificial chromosome into the naturally occurring chromosome of a second plant species. The choice of the specific recombination system used will be dependent on the nature of the modification contemplated.

By having the ability to isolate an artificial chromosome, in particular, artificial chromosomes containing plant chromosomal DNA
10 introduced via site-specific recombination, and re-introduce the chromosome into other mammalian or plant cells, particularly plant cells, these new combinations can be evaluated in different crop species without the need to first isolate and modify the genes, or carry out multiple transformations or gene transfers to achieve the same
15 combination isolation and testing combinations of the genes in plants. The use of a site-specific recombinase also allows the convenient recovery of the plant chromosomal region into other recombinant DNA vectors and systems, such as mammalian or insect systems, for manipulation and study.

20 Also contemplated herein are *ACes*, cell lines and methods for use in screening a new chromosomal combinations, deletions, truncations with eucaryotic genome that take advantage of the site-specific recombination systems incorporated onto platform *ACes* provided herein. For example, provided herein is a cell line useful for making a library of
25 *ACes*, comprising a multiplicity of heterologous recombination sites randomly integrated throughout the endogenous chromosomes. Also provided herein is a method of making a library of *ACes* comprising random portions of a genome, comprising introducing one or more *ACes* into a cell line comprising a multiplicity of heterologous recombination

sites randomly integrated throughout the endogenous chromosomes,
under conditions that promote the site-specific chromosomal arm
exchange of the *ACes* into, and out of, a multiplicity of the heterologous
recombination sites within the cell's chromosomal DNA; and isolating said
5 multiplicity of *ACes*, thereby producing a library of *ACes* whereby multiple
ACes have different portions of the genome within. Also provided herein
is a library of cells useful for genomic screening, said library comprising a
multiplicity of cells, wherein each cell comprises an *ACes* having a
mutually exclusive portion of a chromosomal nucleic acid therein. The
10 library of cells can be from a different species and/or cell type than the
chromosomal nucleic acid within the *ACes*. Also provided is a method of
making one or more cell lines, comprising

- a) integrating into endogenous chromosomal DNA of a selected cell
species, a multiplicity of heterologous recombination sites,
- 15 b) introducing a multiplicity of *ACes* under conditions that promote
the site-specific chromosomal arm exchange of the *ACes* into, and out of,
a multiplicity of the heterologous recombination sites integrated within the
cell's endogenous chromosomal DNA;
- c) isolating said multiplicity of *ACes*, thereby producing a library of
20 *ACes* whereby a multiplicity of *ACes* have mutually exclusive portions of
the endogenous chromosomal DNA therein;
- d) introducing the isolated multiplicity of *ACes* of step c) into a
multiplicity of cells, thereby creating a library of cells;
- e) selecting different cells having mutually exclusive *ACes* therein
25 and clonally expanding or differentiating said different cells into clonal cell
cultures, thereby creating one or more cell lines.

These *ACes*, cell lines and methods utilize the site-specific
recombination sites on platform *ACes* analogous YAC manipulation
related to: the methods of generating terminal deletions in normal and

artificial chromosomes (e.g., *ACes*; as described in Vollrath et al., 1988, *PNAS, USA*, 85:6027-66031; and Pavan et al., *PNAS, USA*, 87:1300-1304); the methods of generating interstitial deletions in normal and artificial chromosomes (as described in Campbell et al., 1991, *PNAS, USA*, 888:5744-5748); and the methods of detecting homologous recombination between two *ACes* (as described in Cellini et al., 1991, *Nuc. Acid Res.*, 19(5):997-1000).

5. Use of platform *ACes* in Pharmacogenomic/toxicology applications (development of "Reporter *ACes*")

10 In addition to the placement of genes onto *ACes* chromosomes for therapeutic protein production or gene therapy, the platform can be engineered via the IntR lambda integrase to carry reporter-linked constructs (reporter genes) that monitor changes in cellular physiology as measured by the particular reporter gene (or a series of different reporter genes) readout.

15 The reporter linked constructs are designed to include a gene that can be detected (by for example fluorescence, drug resistance, immunohistochemistry, or transcript production, and the like) with well-known regulatory sequences that would control the expression of the detectable gene. Exemplary regulatory promoter sequences are well-known

20 in the art.

A) Reporter *ACes* for drug pathway screening

The *ACes* can be engineered to carry reporter-linked constructs that indicate a signal is being transduced through one or a number of pathways. For example, transcriptionally regulated promoters from genes at the end (or

25 any other chosen point) of particular signal transduction pathways could be engineered on the *ACes* to express the appropriate readout (either by fluorescent protein production or drug resistance) when the pathway is activated (or down-regulated as well). In one embodiment, a number of reporters from different pathways can be placed on an

ACes chromosome. Cells (and/or whole animals) containing such a Reporter *ACes* could be exposed to a variety of drugs or compounds and monitored for the effects of the drugs or compounds upon the selected pathway(s) by the reporter gene(s). Thus, drugs or compounds can be
5 classified or identified by particular pathways they excite or down-regulate. Similarly, transcriptional profiles obtained from genomic array experiments can be biologically validated using the reporter *ACes* provided herein.

B) Reporter *ACes* for toxic compound testing

10 Environmental or man-made genotoxicants can be tested in cell lines carrying a number of reporter-genes platform *ACes* linked to promoters that are transcriptionally regulated in response to DNA damage, induced apoptosis or necrosis, and cell-cycle perturbations. Furthermore, new drugs and/or compounds could be tested in a similar manner with the
15 genotoxicant *ACes* reporter for their cellular/genetic toxicity by such a screen. Likewise, toxic compound testing could be carried out in whole transgenic animals carrying the *ACes* chromosome that measures genotoxicant exposure ("canary in a coal mine"). Thus, the same or similar type *ACes* could be used for toxicity testing in either a cell-based
20 or whole animal setting. An example would include *ACes* that carry reporter-linked genes controlled by various cytochrome P450 profiled promoters and the like.

C) Reporter *ACes* for individualized pharmacogenomics/drug profiling

25 A common disease may arise via various mechanisms. In many instances there are multiple treatments available for a given disease. However, the success of a given treatment may depend upon the mechanism by which the disease originated and/or by the genetic background of the patient. In order to establish the most effective

treatment for a given patient one could utilize the *ACes* reporters provided herein. *ACes* reporters can be used in patient cell samples to determine an individualized drug regimen for the patient. In addition, potential polymorphisms affecting the transcriptional regulation of an individual's particular gene can be assessed by this approach.

D) Reporter *ACes* for classification of similar patient tumors

As with other diseases as described in 5.C) above, cancer cells arise via different mechanisms. Furthermore, as a cancerous cell propagates it may undergo genomic alterations. An *ACes* reporter transferred to cells of different patients having the same disease, i.e. similar cancers, could be used to categorize the particular cancer of each patient, thereby facilitating the identification of the most effective therapeutic regimen. Examples would include the validation of array profiling of certain classes of breast cancers. Subsequently, appropriate drug profiling could be carried out as described above.

E) Reporter *ACes* as a "differentiation" sensor

Using the *ACes* reporter as a "differentiation" sensor in stem cells or other progenitor cells in order to enrich by selection (either FACS based screening, drug selection and/or use of suicide gene) for a particular class of differentiated or undifferentiated cells. For example, in one embodiment, this assay could also be used for compound screening for small molecule modifiers of cell differentiation.

F) Whole animal studies with Reporter *ACes*

Finally, with whole-body fluorescence imaging technology (Yang et al. (2000) PNAS 97:12278) any of the above Reporter *ACes* methods could be used in conjunction with whole-body imaging to monitor reporter genes within whole animals without sacrificing the animal. This would allow temporal and spatial analysis of expression patterns under a given set of conditions. The conditions tested may include for example, normal

differentiation of a stem cell, response to drug or compound treatment whether targeted to the diseased tissue or presented systemically, response to genotoxicants, and the like.

The following examples are included for illustrative purposes only
5 and are not intended to limit the scope of the invention.

EXAMPLE 1

pFK161

Cosmid pFK161 (SEQ ID NO: 118) was obtained from Dr. Gyula Hadlaczky and contains a 9 kb *NotI* insert derived from a murine rDNA
10 repeat (see clone 161 described in PCT Application Publication No. WO97/40183 by Hadlaczky *et al.* for a description of this cosmid). This cosmid, referred to as clone 161 contains sequence corresponding to nucleotides 10,232-15,000 in SEQ ID NO. 26. It was produced by inserting fragments of the megachromosome (see, U.S. Patent No.
15 6,077,697 and International PCT application No. WO 97/40183). For example, H1D3, which was deposited at the European Collection of Animal Cell Culture (ECACC) under Accession No. 96040929, is a mouse-hamster hybrid cell line carrying this megachromosome into plasmid pWE15 (Stratagene, La Jolla, California; SEQ ID No. 31) as
20 follows. Half of a 100 μ l low melting point agarose block (mega-plug) containing isolated SATACs was digested with *NotI* overnight at 37°C. Plasmid pWE15 was similarly digested with *NotI* overnight. The mega-plug was then melted and mixed with the digested plasmid, ligation buffer and T4 DNA ligase. Ligation was conducted at 16°C overnight. Bacterial
25 DH5 α cells were transformed with the ligation product and transformed cells were plated onto LB/Amp plates. Fifteen to twenty colonies were grown on each plate for a total of 189 colonies. Plasmid DNA was isolated from colonies that survived growth on LB/Amp medium and analyzed by Southern blot hybridization for the presence of DNA that

hybridized to a pUC19 probe. This screening methodology assured that all clones, even clones lacking an insert but yet containing the pWE15 plasmid, would be detected.

Liquid cultures of all 189 transformants were used to generate
5 cosmid minipreps for analysis of restriction sites within the insert DNA. Six of the original 189 cosmid clones contained an insert. These clones were designated as follows: 28 (~9-kb insert), 30 (~9-kb insert), 60 (~4-kb insert), 113 (~9-kb insert), 157 (~9-kb insert) and 161 (~9-kb insert). Restriction enzyme analysis indicated that three of the clones
10 (113, 157 and 161) contained the same insert.

For sequence analysis the insert of cosmid clone no. 161 was subcloned as follows. To obtain the end fragments of the insert of clone no. 161, the clone was digested with *NotI* and *BamHI* and ligated with *NotI/BamHI*-digested pBluescript KS (Stratagene, La Jolla, California).
15 Two fragments of the insert of clone no. 161 were obtained: a 0.2-kb and a 0.7-kb insert fragment. To subclone the internal fragment of the insert of clone no. 161, the same digest was ligated with *BamHI*-digested pUC19. Three fragments of the insert of clone no. 161 were obtained: a 0.6-kb, a 1.8-kb and a 4.8-kb insert fragment.

20 The insert corresponds to an internal section of the mouse ribosomal RNA gene (rDNA) repeat unit between positions 7551-15670 as set forth in GENBANK accession no. X82564, which is provided as SEQ ID NO. 18. The sequence data obtained for the insert of clone no. 161 is set forth in SEQ ID NOS. 19-25. Specifically, the individual
25 subclones corresponded to the following positions in GENBANK accession no. X82564 (SEQ ID NO:18) and in SEQ ID NOS. 19-25:

5

Subclone	Start	End	Site	SEQ ID No.
	in X82564			
161k1	7579	7755	<i>NotI, BamHI</i>	19
161m5	7756	8494	<i>BamHI</i>	20
161m7	8495	10231	<i>BamHI</i>	21 (shows only sequence corresponding to nt. 8495-8950), 22 (shows only sequence corresponding to nt. 9851- 10231)
161m12	10232	15000	<i>BamHI</i>	23 (shows only sequence corresponding to nt. 10232-10600), 24 (shows only sequence corresponding to nt. 14267-15000)
161k2	15001	15676	<i>NotI, BamHI</i>	25

10 The sequence set forth in SEQ ID NOs. 19-25 diverges in some positions from the sequence presented in positions 7551-15670 of GENBANK accession no. X82564. Such divergence may be attributable to random mutations between repeat units of rDNA.

15 For use herein, the rDNA insert from the clone was prepared by digesting the cosmid with *NotI* and *BglII* and was purified as described above. Growth and maintenance of bacterial stocks and purification of plasmids were performed using standard well known methods (see, *e.g.*, Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory Press), and plasmids were purified from bacterial cultures using Midi - and Maxi-preps Kits (Qiagen, 20 Mississauga, Ontario).

pDsRed1N1

25 This vector is available from Clontech (see SEQ ID No. 29) and encodes the red fluorescent protein (DsRed; Genbank accession no. AF272711; SEQ ID Nos. 39 and 40). DsRed, which has a vivid red fluorescence, was isolated from the IndoPacific sea anemone relative *Discosoma* species. The plasmid pDsRed1N1 (Clontech; SEQ ID No. 29) constitutively expresses a human codon-optimized variant of the

fluorescent protein under control of the CMV promoter. Unmodified, this vector expresses high levels of DsRed1 and includes sites for creating N-terminal fusions by cloning proteins of interest into the multiple cloning site (MCS). It is Kan and Neo resistant for selection in bacterial or
5 eukaryotic cells.

Plasmid pMG

Plasmid pMG (InvivoGen, San Diego, California; see SEQ. ID. NO. 27 for the nucleotide sequence of pMG) contains the hygromycin phosphotransferase gene under the control of the immediate-early human
10 cytomegalovirus (hCMV) enhancer/promoter with intron A. Vector pMG also contains two transcriptional units allowing for the coexpression of two heterologous genes from a single vector sequence.

The first transcriptional unit of pMG contains a multiple cloning site for insertion of a gene of interest, the hygromycin phosphotransferase
15 gene (*hph*) and the immediate-early human cytomegalovirus (hCMV) enhancer/promoter with intron A (see, *e.g.*, Chapman *et al.* (1991) *Nuc. Acids Res.* 19:3979-3986) located upstream of *hph* and the multiple cloning site, which drives the expression of *hph* and any gene of interest inserted into the multiple cloning site as a polycistronic mRNA. The first
20 transcriptional unit also contains a modified EMCV internal ribosomal entry site (IRES) upstream of the *hph* gene but downstream of the hCMV promoter and MCS for ribosomal entry in translation of the *hph* gene (see SEQ ID NO. 27, nucleotides 2736-3308). The IRES is modified by insertion of the constitutive *E. coli* promoter (EM7) within an intron (IM7)
25 into the end of the IRES. In mammalian cells, the *E. coli* promoter is treated as an intron and is spliced out of the transcript. A polyadenylation signal from the bovine growth hormone (bGh) gene (see, *e.g.*, Goodwin and Rottman (1992) *J. Biol. Chem.* 267:16330-16334) and a pause site derived from the 3' flanking region of the human $\alpha 2$

globin gene (see, *e.g.*, Enriquez-Harris *et al.* (1991) *EMBO J.* 10:1833-1842) are located at the end of the first transcription unit. Efficient polyadenylation is facilitated by inserting the flanking sequence of the bGh gene 3' to the standard AAUAAA hexanucleotide sequence.

5 The second transcriptional unit of pMG contains another multiple cloning site for insertion of a gene of interest and an EF-1 α /HTLV hybrid promoter located upstream of this multiple cloning site, which drives the expression of any gene of interest inserted into the multiple cloning site. The hybrid promoter is a modified human elongation factor-1 alpha (EF-1
10 alpha) gene promoter (see, *e.g.*, Kim *et al.* (1990) *Gene* 91:217-223) that includes the R segment and part of the U5 sequence (R-U5') of the human T-cell leukemia virus (HTLV) type I long terminal repeat (see, *e.g.*, Takebe *et al.* (1988) *Mol. Cell. Biol.* 8:466-472). The Simian Virus 40 (SV40) late polyadenylation signal (see Carswell and Alwine (1989) *Mol.*
15 *Cell. Biol.* 9:4248-4258) is located downstream of the multiple cloning site. Vector pMG contains a synthetic polyadenylation site for the first and second transcriptional units at the end of the transcriptional unit based on the rabbit β -globin gene and containing the AATAAA hexanucleotide sequence and a GT/T-rich sequence with 22-23
20 nucleotides between them (see, *e.g.*, Levitt *et al.* (1989) *Genes Dev.* 3:1019-1025). A pause site derived from the C2 complement gene (see, Moreira *et al.* (1995) *EMBO J.* 14:3809-3819) is also located at the 3' end of the second transcriptional unit.

 Vector pMG also contains an ori sequence (ori pMB1) located
25 between the SV40 polyadenylation signal and the synthetic polyadenylation site.

EXAMPLE 2

A. Construction of targeting vector and transfection into LMtk- cells for the generation of platform chromosomes

A targeting vector derived from the vector pWE15 (GeneBank Accession # X65279) was modified by replacing the *SaI* (Klenow filled)/*SmaI* neomycin resistance containing fragment with the *PvuII*/*Bam*HI (Klenow filled) puromycin resistance containing fragment
5 (isolated from plasmid pPUR, Clontech Laboratories, Inc. Palo Alto, CA; SEQ ID No. 30) resulting in plasmid pWEPuro. Subsequently a 9 Kb *NotI* fragment from the plasmid pFK161 (SEQ ID NO: 118) containing a portion of the mouse rDNA region was cloned into the *NotI* site of pWEPuro resulting in plasmid pWEPuro9K (Figure 2). The vector pWEPuro9K was
10 digested with *SpeI* to linearize and transfected into LMtk- mouse cells. Puromycin resistant colonies were isolated and subsequently tested for artificial chromosome formation via fluorescent *in situ* hybridization (FISH) (using mouse major and minor DNA repeat sequences, the puromycin gene and telomeres sequences as probes), and fluorescent activated cell
15 sorting (FACS). From this sort, a subclone was isolated containing an artificial chromosome, designated 5B11.12, which carries 4-8 copies of the puromycin resistance gene contained on the pWEPuro9K vector. FISH analysis of the 5B11.12 subclone demonstrated the presence of telomeres and mouse minor on the ACes. DOT PCR has been done on
20 the 5B11.12 ACes revealing the absence of uncharacterized euchromatic regions on the ACes. A recombination site, such as an *att* or *loxP* engineering site or a plurality thereof, was introduced onto this ACes thereby providing a platform for site-specific introduction of heterologous nucleic acid.

25 **B. Targeting a single sequence specific recombination site onto platform chromosomes**

After the generation of the 5B11.12 platform, a single sequence-specific recombination site is placed onto the platform chromosome via homologous recombination. For this, DNA sequences containing the site-

specific recombination sequence can be flanked with DNA sequences of
homology to the platform chromosome. For example, using the platform
chromosome made from the pWEPuro9K vector, mouse rDNA sequences
or mouse major satellite DNA can be used as homologous sequences to
5 target onto the platform chromosome. A vector is designed to have these
homologous sequences flanking the site-specific recombination site and,
after the appropriate restriction enzyme digest to generate free ends of
homology to the platform chromosome, the DNA is transfected into cells
harboring the platform chromosome (Figure 3). Examples of site-specific
10 cassettes that are targeted to the platform chromosome using either
mouse rDNA or mouse major repeat DNA include the SV40-attP-hygro
cassette and a red fluorescent protein (RFP) gene flanked by loxP sites
(Cre/lox, see, *e.g.*, U.S. Patent No. 4,959,317 and description herein).
After transfection and integration of the site-specific cassette,
15 homologous recombination events onto the platform chromosome are
subcloned and identified by FACS (*e.g.* screen and single cell subclone via
expression of resistance or fluorescent marker) and PCR analysis.

For example, a vector can be constructed containing regions of the
mouse rDNA locus flanking a gene cassette containing the SV40 early
20 reporter-bacteriophage lambda attP site-hygromycin selectable marker
(see Figure 4 and described below). The use of the bacteriophage lambda
attP site for lambda integrase-mediated site-specific recombination is
described below. Homologous recombination event of the SV40-attP-
hygro cassette onto the platform chromosome was identified using PCR
25 primers that detect the homologous recombination and further confirmed
by FISH analysis. After identifying subcloned colonies containing the
platform chromosome with a single site-specific recombination site, cells
carrying the platform chromosome with a single site-specific

recombination site can now be engineered with site-specific recombinases (e.g. lambda INT, Cre) for integrating a target gene expression vector.

C. Targeting a red fluorescent protein (RFP) gene flanked by loxP sites onto 5B11.12 platform

As another example, while loxP recombination sites could have been introduced onto the *ACes* during *de novo* biosynthesis, it was thought that this might result in multiple segments of the *ACes* containing a high number of loxP sites, potentially leading to instability upon *Cre*-mediated recombination. A gene targeting approach was therefore devised to introduce a more limited number of loxP recombination sites into a locus of the 5B11-12 *ACes* containing introduced and possibly co-amplified endogenous rDNA sequences. Although there are more than 200 copies of rDNA genes in the haploid mouse genome distributed amongst 5-11 chromosomes (depending on strain), rDNA sequences were chosen as the target on the *ACes* since they represent a less frequent target than that of the satellite repeat sequences. Moreover, having observed much stronger pWEPuro9K hybridization to the 5B11-12 *ACes* than to other LMTK chromosomes and in light of the observation that the transcribed spacer sequences within the rDNA may be less conserved than the rRNA coding regions, it was contemplated that a targeting vector based on the rDNA gene segment in pWEPuro9K would have a higher probability of targeting to the *ACes* rather than to other LMTK chromosomes. Accordingly, a targeting vector, pBSFKLoxDsRedLox, was designed and constructed based on the rDNA sequences contained in pWEPuro9K.

The plasmid pBSFKLoxDsRedLox was generated in 4 steps. First, the *Nox1* rDNA insert of pWEPuro9K (Figure 2) was inserted into pBS SK- (Stratagene) giving rise to pBSFK. Second, a loxP polylinker cassette was

generated by PCR amplification of pNEB193 (SEQ ID NO:32; New England Biolabs) using primers complementary to the M13 forward and reverse priming sites at their 3' end and a 34 bp 5' extension comprising a LoxP site. This cassette was reinserted into pNEB193 generating

5 p193LoxMCSLox. Third, the DsRed gene from pDsRed1-N1 (SEQ ID NO:29; Clontech) was then cloned into the polylinker between the loxP sites generating p193LoxDsRedLox. Fourth, a fragment consisting of the DsRed gene flanked by loxP sites was cloned into a unique *NdeI* within the rDNA insert of pBSFK generating pBSFKLoxDsRedLox.

10 A gel purified 11 Kb *PmlI* / *EcoRV* fragment of pBSFKLoxDsRedLox was used for transfection. To detect targeted integration, PCR primers were designed from rDNA sequences within the 5' *NotI*-*PmlI* fragment of pWEPuro9K that is not present on the targeting fragment (5' primer) and sequence within the LoxDsRedLox cassette (3' primer). If the targeting
15 DNA integrated correctly within the rDNA sequences, PCR amplification using these primers would give rise to a 2.3 Kb band. PCR reactions containing 1-4 μ l of genomic DNA were carried out according to the MasterTaq protocol (Eppendorf), using murine rDNA 5' primer (5'-CGGACAATGCGGTTGTGCGT-3'; SEQ ID NO:72) and DsRed 3' primer
20 (5'-GGCCCCGTAATGCAGAAGAA-3'; SEQ ID NO:73) and PCR products were analyzed by agarose gel electrophoresis.

1.5X10⁶ 5B11-12 LMTK⁻ cells were transfected with 2 μ g of the pBSFKLoxDsRedLox targeting DNA described above using Lipofectamine Plus (Invitrogen). For flow sorting, harvested cells were suspended in
25 medium and applied to the Becton Dickinson Vantage SE cell sorter, equipped with 488 nm lasers for excitation and 585/42 bandpass filter for optimum detection of RFP fluorescence. Cells were sorted using dPBS as sheath buffer. Negative control parental 5B11-12 cells and a positive control LMTK⁻ cell line stably transfected with DsRed were used to

establish the selection gates. The RFP positive gated populations were recovered, diluted in medium supplemented with 1X penicillin-streptomycin (Invitrogen), then plated and cultured as previously described. After 4 rounds of enrichment, the percentage of RFP positive cells reached levels of 50% or higher. DNA from populations was analyzed by PCR for evidence of targeted integration. Ultimately, single cell subclones were established from positive pools and were analyzed by PCR and PCR-positive clones confirmed by FISH as described below. DNA was purified from pools or single cell clones using previously described methods set forth in Lahm et al., Transgenic Res., 1998; 7:131-134, or in some cases using a Wizard Genomic DNA purification kit (Promega). For FISH analysis, a biotinylated DsRed gene probe was generated by PCR using DsRed specific primers and biotin-labeled dUTP (5' RFP primer: 5'-GGTTTAAAGTGCCTCCTCCAAGAACGTCATC-3', SEQ ID NO:74; and 3' RFP primer: 5'AGATCTAGAGCCGCGCTACAGGAACAGGTGGTGGCGGCC-3'; SEQ ID NO:75). To maximize the signal intensity of the DsRed probe, Tyramide amplification was carried out according to the manufacturers protocols (NEN).

20 The process of testing the feasibility of a more general targeting strategy that would not rely on enrichment *via* drug selection of stably transfected clones can be summarized as follows. A red fluorescent protein gene (RFP; encoded by the DsRed gene) was inserted between the loxP sites of the targeting vector to form pBSFKLoxDsRedLox. After

25 transfection with pBSFKLoxDsRedLox, sequential rounds of high speed flow sorting and expansion of sorted cells in culture could then be used to enrich for stable transformants expressing RFP. In the event of targeted integration, PCR screening with primers that amplify from a spacer region within the segment of the 45s pre-rRNA gene in pWEPuro9K to a specific

anchor sequence within the DsRed gene in the targeting cassette would give rise to a diagnostic 2.3 Kb band. However, as rDNA clusters are found on several chromosomes, confirmation of targeting to an *ACes* would require fluorescence in situ hybridization (FISH) analysis. Finally, the flanking of the DsRed gene by loxP sites would allow for its removal and subsequent replacement with other genes of interest.

After transfection of the targeting sequence into 5B11-12 cells, enrichment for targeted clones was carried out using a combination of flow cytometry to detect red-fluorescing cells and PCR screening.

- 10 Ultimately 17 single cell subclones were identified as potential targeted clones by PCR and of these 16 were found by FISH to contain the DsRed integration event into the *ACes*. These subclones are referred to herein as D11-C4, D11-C12, D11-H3, C9-C9, C9-B9, C9-F4, C9-H8, C9-F2, C9-G8, C9-B6, C9-G3, C9-E12, C9-A11, C11-E3, C11-A9 and C11-H4. PCR
- 15 analysis of genomic DNA isolated from the D11-C4 subclone gave rise to a 2.3 Kb band, indicative of a targeted integration into an rDNA locus. Further analysis of the subclone by FISH analysis with a DsRed gene probe demonstrated integration of the LoxDsRedLox targeting cassette on the *ACes* co-localizing with one of the regions of rDNA staining seen on
- 20 the 5B11-12 *ACes*, consistent with a targeted integration into an rDNA locus of the *ACes*, while integrations on other chromosomes were not observed. Since transfected cells were maintained as heterogeneous populations through several cycles of sorting and replating it was not possible to estimate the frequency of targeted events. In most
- 25 mammalian cell lines the frequency of gene targeting via homologous recombination is roughly 10^{-5} - 10^{-7} treated cells. Despite the low frequency of these events in mammalian cells, it is clear that an RFP expression based screening paradigm, coupled with PCR analysis, can effectively detect and enrich for such infrequent events in a large

population. In instances where drug selection is not possible or not desirable, such a system may provide a useful alternative. It was also verified that the modified *ACes* in subclone D11-C4 could be purified by flow cytometry. The results indicate that the flow karyogram of the D11-C4 subclone was unaltered from that of the 5B11-12 cell line. Thus, the D11-C4 *ACes* can be purified in high yield from native chromosomes of the host cell line.

D. Reduction of LoxP on *ACes* to a single site.

The strong hybridization signal detected by FISH on the *ACes* using the DsRed gene probe suggests that several copies of the targeting cassette may be present on the *ACes* in the D11-C4 line. This also suggests that multiple rDNA genes have been correctly targeted.

Accordingly, in certain embodiments where necessary, the number of loxP sites on the *ACes* can be reduced to a single site by *in situ* treatment with *Cre* recombinase, provided that the sites are co-linear. Such a process is described for multiple loxP-flanked integrations on a native mouse chromosome (Garrrick et al., Nature Genet., 1998, Jan;18(1):56-59). Reduction to a single loxP site on the D11-C4 *ACes* would result in the loss of the DsRed gene, forming the basis of a useful screen for this event.

For this purpose, a Cre expression plasmid pCX-Cre/GFP III has been generated by first deleting the EcoRI fragment of pCX-eGFP (SEQ ID NO:71) containing the eGFP coding sequence and replacing it with that of a PCR amplified Cre recombinase coding sequence (SEQ ID NO:58), generating pCX-Cre. Next, the AseI/SspI fragment of pD2eGFP-N1 (containing the CMV promoter driving the D2EGFP gene with SV40 polyA signal; Clontech; SEQ ID NO:87) was inserted into the filled HindIII site of pCX-Cre, generating pCX-Cre\GFP III. Control plasmid pCX-CreRev\GFP

III was generated in similar fashion except that the Cre recombinase coding sequence was inserted in the antisense orientation. LMTK⁻ cell line D11-C4 (containing first generation platform ACes with multiple loxP-DsRED sites) and 5B11-12 cell line (containing ACes with no loxP-DsRED sites) are maintained in culture as described above. D11C4 cells are transfected with 2 μ g of plasmid pCX-Cre\GFP III or 2 μ g pCX-CreRev\GFP III using Lipofectamine (Invitrogen) as previously described.

Forty-eight to seventy-two hours after transfection, transfected D11-C4 cells are harvested and GFP positive cells are sorted by cell cytometry using a FACSta Vantage cell sorter (Beckton-Dickinson) as follows: All D11-C4 cells transfected with pCX-Cre\GFP III or control plasmid pCX-CreRev\GFP III that exhibit GFP fluorescent higher than the gate level established by untransfected cells are collected and placed in culture a further 7-14 days. After 7-14 days the initial D11-C4 cells are harvested and analyzed by cell cytometry as follows: Untransfected D11-C4 cells are used to establish the gate that defines the RFP positive population, while 5B11-12 cells are used to set the RFP negative gate. The GFP positive population of D11-C4 transfected with pCX-Cre\GFP III should show decreased red fluorescence compared to pCX-CreRev\GFP III transfected or untransfected control D11-C4 cells. The cells exhibiting greatly decreased or no RFP expression are collected and single cell clones subsequently established. These clones will be expanded and analyzed by fluorescence *in-situ* hybridization and Southern blotting to confirm the removal of loxP-DsRed gene copies.

25

EXAMPLE 3

Construction of targeting vector and transfection into LMtk⁻ cells for the generation of platform chromosomes containing multiple site-specific recombination sites

An example of a selectable marker system for the creation of a chromosome-based platform is shown in Figure 4. This system includes a vector containing the SV40 early promoter immediately followed by (1) a 282 base pair (bp) sequence containing the bacteriophage lambda attP site and (2) the puromycin resistance marker. Initially a *PvuII/StuI* fragment containing the SV40 early promoter from plasmid pPUR (Clontech Laboratories, Inc., Palo Alto, CA; Seq ID No. 30) was subcloned into the *EcoRI/CRI* site of pNEB193 (a PUC19 derivative obtained from New England Biolabs, Beverly, MA; SEQ ID No. 32) generating the plasmid pSV40193. The only differences between pUC19 and pNEB193 are in the polylinker region. A unique *Ascl* site (GGCGCGCC) is located between the *BamHI* site and the *SmaI* site, a unique *PacI* site (TTAATTAA) is located between the *BamHI* site and the *XbaI* site and a unique *PmeI* site (GTTTAAAC) is located between the *PstI* site and the *SaI* site.

The attP site was PCR amplified from lambda genome (GenBank Accession # NC 001416) using the following primers:

attPUP: CCTTGCGCTAATGCTCTGTTACAGG SEQ ID No. 1

attPDWN: CAGAGGCAGGGAGTGGGACAAAATTG SEQ ID No. 2

After amplification and purification of the resulting fragment, the attP site was cloned into the *SmaI* site of pSV40193 and the orientation of the attP site was determined by DNA sequence analysis (plasmid pSV40193attP). The gene encoding puromycin resistance (Puro) was isolated by digesting the plasmid pPUR (Clontech Laboratories, Inc. Palo Alto, CA) with *AgeI/BamHI* followed by filling in the overhangs with Klenow and subsequently cloned into the *Ascl* site downstream of the attP site of pSV40193attP generating the plasmid pSV40193attPsensePUR (Figure 4; SEQ ID NO:113)).

The plasmid pSV40193attPsensePUR was digested with *ScaI* and co-transfected with the plasmid pFK161 (SEQ ID NO: 118) into mouse LMtk- cells and platform artificial chromosomes were identified and isolated as described above. The process for generating this exemplary platform ACes containing multiple site-specific recombination sites is summarized in Figure 5. One platform ACes resulting from this experiment is designated B19-18. This platform ACes chromosome may subsequently be engineered to contain target gene expression nucleic acids using the lambda integrase mediated site-specific recombination system as described herein in Example 7 and 8.

EXAMPLE 4

Lambda integrase mediated site-specific recombination of a RFP expressing vector onto artificial chromosomes

In this example, a vector expressing the red fluorescent protein (RFP) was produced and recombined into the attP site residing on an artificial chromosome within LMTK- cells. This recombination is depicted in Figure 7.

A. Construction of expression vectors containing wildtype and mutant lambda integrase

Mutations at the glutamic acid at position 174 in the lambda integrase protein relaxes the requirement for the accessory protein IHF during recombination and DNA supercoiling *in vitro* (see, Miller *et al.* (1980) *Cell* 20:721-729; Lange-Gustafson *et al.* (1984) *J. Biol. Chem.* 259:12724-12732). Mutations at this site promote attP, attB intramolecular recombination in mammalian cells (Lorbach *et al.* (2000) *J. Mol. Biol* 296:1175-1181).

To construct nucleic acid encoding the mutant, lambda integrase was PCR amplified from bacteriophage lambda DNA (cl857 *ind Sam 7*; New England Biolabs) using the following primers:
Lamint1 (SEQ ID No. 3)

TTCGAATTCATGGGAAGAAGGCGAAGTCATGAGCG)

Lamint2 (SEQ ID No. 4)

(TTCGAATTCTTATTTGATTTCAATTTTGTCCCAC).

The resulting PCR product was digested with *EcoR* I and cloned into the
5 *EcoR* I site of pUC19. Lambda integrase was mutated at amino acid
position 174 using QuikChange Site-Directed Mutagenesis Kit
(Stratagene) and the following oligos (generating a glutamic acid to
arginine change at position 174):

LambdaINTE174R

10 (SEQ ID No. 6)

(CGCGCAGCAAAATCTAGAGTAAGGAGATCAAGACTTACGGCTGACG),

LamintR174rev (SEQ ID No. 7)

(CGTCAGCCGTAAGTCTTGATCTCCTTACTCTAGATTTTGCTGCGCG).

The resulting site directed mutant was confirmed by sequence analysis.

15 The wildtype and mutant lambda genes were cloned into the *EcoR* I site
of pCX creating pCX-LamInt (SEQ ID NO: 127) and pCXLamIntR (Figure
8; SEQ ID NO: 112).

The plasmid pCX (SEQ ID No. 70) was derived from plasmid
pCXeGFP (SEQ ID No. 71). Excision of the *EcoRI* fragment containing the
20 eGFP marker generated pCX. To generate plasmid pCXLamINTR (SEQ ID
NO: 112) an *EcoRI* fragment containing the lambda integrase E174R (SEQ
ID No. 37) mutation was cloned into the *EcoRI* site of pCX, and to
generate plasmid pCX-LamINT, an *EcoRI* fragment containing the wild-
type lambda integrase was cloned into the *EcoRI* site of pCX.

25 **B. Construction of integration vector containing attB and DsRed**

The plasmid pDsRedN1 (Clontech Laboratories, Palo Alto, CA; SEQ
ID No. 29) was digested with *Hpa* I and ligated to the following annealed
oligos:

attB1 (SEQ ID No. 8)

(TGAAGCCTGCTTTTTTATACTAACTTGAGCGAA)

attB2 (SEQ ID No. 9)

(TTCGCTCAAGTTAGTATAAAAAAGCAGGCTTCA)

The resulting vector (pDsRedN1-attB) was confirmed by PCR and

5 sequence analysis.

C. Transfection into LMtk- cells

LM(tk-) cells containing the Prototype A ACes (L1-18; Chromos
Molecular Systems Inc., Burnaby, BC Canada) were co-transfected with
pDsRedN1 or pDsRedN1-attB and either pCXLamInt (SEQ ID NO: 127) or
10 pCXLamIntR (SEQ ID NO: 112) using Lipofectamine Plus Reagent
(LifeTechnologies, Gaithersburg, MD). The transfected cells were grown
in DMEM (LifeTechnologies, Gaithersburg, MD) with 10% FBS (CanSera)
and G418 (CalBiochem) at a concentration of 1 mg/ml.

D. Enrichment by cell sorting

15 The transfected cells were sorted using a FACs Vantage SE cell
sorter (Becton Dickenson) to enrich for cells expressing DsRed. The cells
were excited with a 488 nm Argon laser at 200 watts and cells
fluorescing in the 585/42 detection channel were collected. The sorted
cells were returned to growth medium for recovery and expansion. After
20 three successive enrichments for cells expressing DsRed, single cell
sorting into 96 well plates was performed using the same parameters.
Duplicate plates of the single cell clones were made for PCR analysis.

E. PCR analysis of single cell clones

25 Pools of cells from each row and column of the 96 well plate were
used for DNA isolation. DNA was prepared using a Wizard Genomic DNA
purification kit (Promega Inc, Madison, WI). Nested PCR analysis on the
DNA pools was performed to confirm the site-specific recombination
event using the following primer sets:

attPdwn2 (SEQ ID No. 10)
(TCTTCTCGGGCATAAGTCGGACACC)

CMVen (SEQ ID No. 11)
(CTCACGGGGATTCCAAGTCTCCAC)

5 followed by:

attPdwn (SEQ ID No. 12)
(CAGAGGCAGGGAGTGGGACAAAATTG)

CMVen2 (SEQ ID No. 13)
(CAACTCCGCCCCATTGACGCAAATG).

10 The resulting PCR reactions were analyzed by gel electrophoresis and the potential individual clones containing the site-specific recombination event were identified by combining the PCR results of all of the pooled rows and columns for each 96 well plate. The individual clones were then further analyzed by PCR using the following primers that flank the recombination junction. L1for and F1rev flank the attR junction whereas REDfor and L2rev flank the attL junction (see Figure 7):

L1for (SEQ ID No. 14)
AGTATCGCCGAACGATTAGCTCTTCA

20 F1rev (SEQ ID No. 15)
GCCGATTTCGGCCTATTGGTTAAA

REDfor (SEQ ID No. 16)
CCGCCGACATCCCCGACTACAAGAA

L2rev (SEQ ID No. 17)
TTCCTTCGAAGGGGATCCGCCTACC.

25 **F. Sequence analysis of recombination junctions**

PCR products spanning the recombination junction were Topo-cloned into pcDNA3.1D/V5His (Invitrogen Inc., San Diego, CA) and then sequenced by cycle-sequencing. The clones were confirmed to have the correct *attR* and *attL* junctions by cycle sequencing.

30 **G. Fluorescent In Situ Hybridization (FISH)**

The cell lines containing the correct recombination junction sequence were further analyzed by fluorescent *in situ* hybridization (FISH)

by probing with the DsRed coding region labeled with biotin and visualizing with the Tyramide Signal Amplification system (TSA; NEN Life Science Products). The results indicate that the RFP sequence is present on the *ACes*.

5 **H. Southern analysis**

Genomic DNA was harvested from the cell lines containing an *ACes* with the correct recombinant event and digested with *EcoR* I. The digested DNAs were separated on a 0.7% agarose gel, transferred and fixed to a nylon membrane and probed with RFP coding sequences. The
10 result showed that there is an integrated copy of RFP coding sequence in each clone.

EXAMPLE 5

Delivery of a second gene encoding GFP onto the RFP platform *ACes*

15 **A. Construction of integration vector containing attB and GFP (pD2eGFPIresPuroattB).**

The plasmid pIRESpuo2 (Clontech, Palo Alto, CA; SEQ ID NO: 88) was digested with *EcoRI* and *NotI* then ligated to the D2eGFP *EcoRI-NotI* fragment from pD2eGFP-N1 (Clontech, Palo Alto, CA) to create pD2eGFPIresPuro2. Subsequently, oligos encoding the attB site were
20 annealed and ligated into the *NruI* site of pD2eGFPIresPuro2 to create pD2eGFPIresPuroattB. The orientation of attB in the *NruI* site was determined by PCR.

B. Transfection of LMtk- cells

The LMtk- cells containing the RFP platform *ACes* produced in
25 Example 4, which has multiple attP sites, were co-transfected with pCXLaIntR and pD2eGFPIresPuroattB using LipofectAMINE PLUS reagent. Five μ g of each vector was placed into a tube containing 750 μ l of DMEM (Dulbecco's modified Eagles Medium). Twenty μ l of the Plus reagent was added to the DNA and incubated at room temperature for 15

minutes. A mixture of 30 μ l of lipofectamine and 750 μ l DMEM was added to the DNA mixture and incubated an additional 15 minutes at room temperature. The DNA mixture was then added dropwise to approximately 3 million cells attached to a 10cm dish in 5 mls of DMEM.

- 5 The cells were incubated 4 hours (37°C, 5% CO₂) with the DNA-lipid mixture, after which DMEM with 20% fetal bovine serum was added to the dishes to bring the culture medium to 10% fetal bovine serum. The dishes were incubated at 37°C with 5% CO₂.

- Plasmid pD2eGFPIresPuroattB has a puromycin gene transcriptionally linked to the GFP gene *via* an IRES element. Two days after the transfection the cells were placed in medium containing puromycin at 4 μ g/ml to select for cells containing the pD2eGFPIresPuroattB plasmid integrated into the genome. Twenty-three clones were isolated after 17 days of selection with puromycin. These clones were expanded and then analyzed for the presence of the GFP gene on the ACes by 2-color (RFP/biotin & GFP/digoxigenin) TSA-FISH (NEN) according to the manufacturers protocol. Sixteen of the 23 clones produced a positive FISH signal on the ACes with a GFP probe.

EXAMPLE 6

20 Delivery Of ACes Into human Mesenchymal Stem Cells (hMSC)

A. Transfection

- Transfection conditions for the most efficient delivery of the ACes into hMSCs (Cambrex BioWhittaker Product Code PT-2501, lot# F0658, East Rutherford, New Jersey) were assayed using LipofectAMINE PLUS and Superfect. One million prototype B ACes, which is a murine derived 60Mb ACes having primarily murine pericentric heterochromatin, and carrying a "payload" containing a hygromycin B selectable marker gene and a *lacZ* reporter gene (see , Telenius et al., 1999, Chrom. Res., 7:3-7; and Kereso et al., 1996, Chrom. Res., 4:226-239; each of which is

incorporated herein by reference in its entirety), were combined with 1-12 μ l of the transfection agent. In the case of LipofectAMINE PLUS, the PLUS reagent was combined with the ACes for 15 minutes followed by LipofectAMINE for a further 15 minutes. Superfect was complexed for 5 10 minutes at a ratio of 2 μ l Superfect per 1 million ACes. The ACes/transfection agent complex was then applied to 0.5 million recipient cells and the transfection was allowed to proceed according to the manufacturer's protocol. Percent transfected cells was determined on a FACS Vantage flow cytometer with argon laser tuned to 488 nm at 10 200mW and FITC fluorescence collected through a standard FITC 530/30 nm band pass filter. After 24 hours, IdUrd labeled ACes were delivered to human MSCs in the range of 30-50%, varying with transfection agent and dose. ACes delivery curves were generated from data collected in experiments that varied the dose of the transfection reagents. Dose 15 response curves of Superfect and LipofectAMINE PLUS, showing delivery of ACes into recipient hMSCs cells, were prepared, measured by transfer of IdUrd labeled ACes and detected by flow cytometry. Superfect shows maximum delivery in the range of 30-50% at doses greater than 2 μ l per million ACes. LipofectAMINE PLUS has a 42-48% delivery peak around 20 5-8 μ l per million ACes. These dose curves were then correlated with toxicity data to determine the transfection conditions that will allow for highest potential transfection efficiency. Toxicity was determined by a modified plating efficiency assay (de Jong et al., 2001, Chrom. Research, 9:475-485). The population's normalized plating efficiency (at maximum 25 % delivery doses) was in the range of 0.2 - 0.4 for Superfect and 0.5 - 0.6 with LipofectAMINE PLUS.

Due to the transfected population consisting of mixed cell types, flow cytometry allowed for the assessment of ACes delivery into each sub-population and the purification of the target population. Flow profiles

showing forward scatter (cell size) and side scatter (internal cell granularity) revealed three distinct hMSC populations that were gated into three regions: R3 (small cell region), R4 (medium cell region), R5 (large cell region). Transfection conditions were further optimized by re-analyzing delivery curves and assessing the differences in delivery to each sub-population. Dose response curves of Superfect and LipofectAMINE were prepared showing % delivery to each sub-population represented by the gating on basis of cell size and granularity properties of the mixed population. Three distinct hMSC populations were gated and % delivery dose curves generated. Using Superfect and LipofectAMINE PLUS the overall % delivery increased with cell size (80-90% delivery in large cells). LipofectAMINE PLUS at high doses (8-12 μ l per 1 million ACes) shows an increase in the overall proportion of chromosome transfer to the small population (10-20%). This suggests an advantage to using this transfection agent if the small-undifferentiated cell population is the desired target host cell.

B. Expression from Genes on ACes IN hMSCs

Following the delivery screening process conducted in section (A) above, the most promising results were subjected to further analyses to monitor expression and verify the presence of structurally intact ACes. The transfection conditions employed for these experiments were exactly the same as those that had been used during the screening process. Short-term expression was monitored by transfecting hMSCs with ACes containing a RFP gene (red fluorescent protein) set forth in Example 2C as "D11C4". The unselected population was harvested at 72-96 hours post transfection and % positive fluorescent cells measured by flow cytometry. RFP expression was in the range of 1-20%.

Long term-gene expression was assayed by selecting for hygromycin B resistant cells over a period of 7-10 days. Cytogenetic

analysis was done to detect presence of intact *ACes* by Fluorescent *In Situ* hybridization (FISH), where metaphase chromosomes were hybridized to a mouse major satellite-DNA probe (targeting murine pericentric heterochromatin) and a lambda probe (hybridizing to the *lacZ* gene). The human mesenchymal transfected culture could not undergo standard sub-cloning as diffuse colonies form with limited doublings available for expansion. Cytogenetic analysis was performed on the entire population, sampling over a period of 3-10 days post-transfection. The hygromycin resistant population was then blocked in mitosis with colchicine and analyzed for presence of intact *ACes* by FISH. Preliminary FISH results show approximately 2-8% of the hMSC-transfected population had an intact *ACes*. This compared to rat skeletal muscle myoblast clones, which were in the range of 60-95%. To increase the % of intact *ACes* in the hMSC-transfected population an enrichment step can be utilized as described in Example 2C.

C. Differentiation of The hMSCs

In initial experiments where transfected hMSCs cells have been induced to differentiate into adipose or osteocytes, the results indicate that the transfected cells appear to be differentiating at a rate comparable to the untransfected controls and the cultures are lineage specific as tested by microscopic examination, FISH, Oil Red O staining (adipocyte assay), and calcium secretion (osteocyte assay).

Accordingly, these results indicate that the artificial chromosomes (*ACes*) provided herein can be successfully transferred into hMSC target cells. Targeting MSCs (such as hMSCs) permits gene transfer into cells in an undifferentiated state where the cells are easier to expand and purify. The genetically modified cells can then be differentiated *in vitro* or injected into a site *in vivo* where the microenvironment will induce transformation into specific cell lineages.

EXAMPLE 7

Delivery of a Promoterless Marker Gene to a Platform ACes

Platform ACes containing pSV40attPsensePURO (Figure 4) were constructed as set forth in Examples 3 and 4.

5 **A. Construction of Targeting Vectors.**

The base vector p18attBZeo (3166bp; SEQ ID NO: 114) was constructed by ligating the 1067bp *HindIII-SspI* fragment containing attBZeo, obtained from pLITattBZeo (SEQ ID NO:91), into pUC18 (SEQ ID NO: 122) digested with *HindIII* and *SspI*.

10 1. p18attBZEO-eGFP (6119bp; SEQ ID NO: 126) was constructed by inserting the 2977bp *SpeI-HindIII* fragment from pCXeGFP (SEQ ID NO:71; Okabe, *et al.* (1997) *FEBS Lett* 407:313-319) containing the eGFP gene into p18attBZeo (SEQ ID NO: 114) digested with *HindIII* and *XbaI*.

15 2. p18attBZEO-5'6XHS4eGFP (Figure 10; 7631bp; SEQ ID NO: 116) was constructed by ligating the 4465bp *HindIII* fragment from pCXeGFPattB(6XHS4)2 (SEQ ID NO: 123), which contains the eGFP gene under the regulation of the chicken beta actin promoter, 6 copies of the HS4 core element located 5' of the chicken beta actin promoter and the polyadenylation signal, into the *HindIII* site of p18attBZeo (SEQ ID NO: 20 114).

25 3. p18attBZEO-3'6XHS4eGFP (Figure 11; 7600bp; SEQ ID NO: 115) was created by removing the 5'6XHS4 element from p18attBZeo-(6XHS4)2eGFP (SEQ ID NO: 110). p18attBZeo-(6XHS4)2eGFP was digested with *EcoRV* and *SpeI*, treated with Klenow and religated to form p18attBZeo3'6XHS4eGFP (SEQ ID NO: 115).

4. p18attBZEO-(6XHS4)2eGFP (Figure 12; 9080bp; SEQ ID NO: 110) was created in two steps. First, the *EcoRI-SpeI* fragment from pCXeGFPattB(6XHS4)2 (SEQ ID NO: 123), which contains 6 copies of the HS4 core element, was ligated into p18attBZeo (SEQ ID NO: 114)

digested with *EcoRI* and *XbaI* to create p18attBZeo6XHS4 (4615bp; SEQ ID NO: 117). Next, p18attBZeo6XHS4 was digested with *HindIII* and ligated to the 4465bp *HindIII* fragment from pCXeGFPattB(6XHS4)2 which contains the eGFP gene under the regulation of the chicken beta actin promoter, 6 copies of the HS4 core element located 5' of the chicken beta actin promoter and the polyadenylation signal.

Table 2

Targeting plasmid	No. zeocin resistant clones	No. clones with expected PCR product size	No. clones with correct sequence at recombination junction
p18attBZEO-eGFP	12	12	NT*
p18attBZEO-5'6XHS4eGFP	11	11	NT
p18attBZEO-3'6XHS4eGFP	11	11	NT
p18attBZEO-(6XHS4)2eGFP	9	9	4/4

*NT = not tested

B. Transfection and Selection with Drug.

The mouse cell line containing the 2nd generation platform ACE, B19-38 (constructed as set forth in Example 3), was plated onto four 10cm dishes at approximately 5 million cells per dish. The cells were incubated overnight in DMEM with 10% fetal calf serum at 37°C and 5% CO₂. The following day the cells were transfected with 5µg of each of the 4 vectors listed in Example 7.A. above and 5µg of pCXLamIntR (SEQ ID NO: 112), for a total of 10µg per 10cm dish. Lipofectamine Plus reagent was used to transfect the cells according to the manufacturers protocol. Two days post-transfection zeocin was added to the medium at 500µg/ml. The cells were maintained in selective medium until colonies formed. The colonies were then ring-cloned (see, e.g., McFarland, 2000, Methods Cell Sci, Mar;22(1):63-66).

C. Analysis of Clones (PCR, SEQUENCING).

Genomic DNA was isolated from each of the candidate clones with the Wizard kit (Promega) and following the manufacturers protocol. The following primer set was used to analyze the genomic DNA isolated from the zeocin resistant clones: 5PacSV40 –

- 5 CTGTTAATTAAGTGTGGAATGTGTG TCAGTTAGGGTG (SEQ ID NO:76);
Antisense Zeo - TGAACAGGGTCACGTCGTCC (SEQ ID NO:77). PCR amplification with the above primers and genomic DNA from the site-specific integration of any of the 4 zeocin vectors would result in a 673bp PCR product.

- 10 As set forth in Table 2, of the 4 zeocin resistant candidate clones thusfar analyzed by PCR, all 4 exhibit the correct sequence for a site-specific integration event.

EXAMPLE 8

Integration of a PCR product by site-specific recombination.

- 15 In this example a gene is integrated onto the platform ACes by site-specific recombination without cloning said gene into a vector.

A. PCR PRIMER DESIGN.

- 20 PCR primers are designed to contain an attB site at the 5' end of one of the primers in the primer set. The remaining primers, which could be one or more than one primer, do not contain an attB site, but are complementary to sequences flanking the gene or genes of interest and any associated regulatory sequences. In first example, 2 primers (one containing an attB site) are used to amplify a selective gene such as puromycin.

- 25 In a second example as shown in Figure 13, the primer set includes primers 1 & 2 that amplify the GFP gene without amplification of an upstream promoter. Primer 1 contains the attB site at the 5' end of the oligo. Primers 3 & 4 are designed to amplify the IRES-blasticidin DNA sequences from the vector pIRESblasticidin. The 5'end of primer 3

contains sequences complementary to the 5' end of primer 2 such that annealing can occur between 5' ends of the two primers.

B. PCR REACTION AND SUBSEQUENT LIGATION TO CREATE CIRCULAR MOLECULES FROM THE PCR PRODUCT

5 In the first example set forth above in Section A, the two PCR primers are combined with a puromycin DNA template such as pPUR (Clontech), a heat stable DNA polymerase and appropriate conditions for DNA amplification. The resulting PCR product (attB-Puromycin) is then then purified and self-ligated to form a circular molecule.

10 In the second example set forth above in Section A, amplification of the GFP gene and IRES-blasticidin sequences is accomplished by combining primers 1 & 2 with DNA template pD2eGFP and primers 3 & 4 with template pIRESblasticidin under appropriate conditions to amplify the desired template. After initial amplification of the two products (attB-GFP
15 & IRES-blasticidin) in separate reactions, a second round of amplification using both of the PCR products from the first round of amplification together with primers 1 and 4 amplifies the fusion product attB-GFP-IRES-blasticidin (Figure 13). This technique of using complementary sequences in primer design to create a fusion product is employed in *Saccharomyces*
20 *cerevisiae* for allele replacement (Erdeniz *et al* (1997) *Gen Res* 7:1174-1183). The amplified product is then purified from the PCR reaction mixture by standard methods and ligated to form a circular molecule.

C. INTRODUCTION OF PCR PRODUCT ONTO THE ACes USING A RECOMBINASE

25 The circular PCR product is then be introduced to the platform ACes using the bacteriophage lambda integrase E174R. The introduction can be performed *in vivo* by transfecting the pCXLamIntR (SEQ ID NO: 112) vector encoding the lambda integrase mutant E174R together with the circularized PCR product into a cell line containing the platform ACE.

D. SELECTION FOR MARKER GENE

The marker gene (in this case either puromycin, blasticidin or GFP) is used to enrich the population for cells containing the proper integration event. A proper integration event in the second example (Figure 14)

- 5 juxtaposes a promoter residing on the platform *ACes* 5' to the attB-GFP-IRES-Blasticidin PCR product, allowing for transcription of both GFP and blasticidin. If enrichment is done by drug selection, blasticidin is added to the medium on the transfected cells 24-48 hours post-transfection. Selection is maintained until colonies are formed on the plates. If
10 enrichment is done by cell sorting, cells are sorted 2-4 days post-transfection to enrich for cells expressing the fluorescent marker (GFP in this case).

E. ANALYSIS OF CLONES

- Clonal isolates are analyzed by PCR, FISH and sequence analysis to
15 confirm proper integration events.

EXAMPLE 9

Construction of a human platform *ACes* "ACE 0.1"

A. CONSTRUCTION OF THE TARGETING VECTOR pPACrDNA

- Genome Systems (IncyteGenomics) was supplied with the primers
20 5'HETS (GGGCCGAAACGATCTCAACCTATT; SEQ ID NO:78), and
3'HETS (CGCAGCGGCCCTCCTACTC; SEQ ID NO:79), which were used to amplify a 538bp PCR product homologous to nt 9680-10218 of the human rDNA sequences (GenBank Accession No. U13369) and used as a probe to screen a human genomic P1AC (P1 Artificial Chromosome)
25 library constructed in the vector pCYPAC2 (Ioannou *et al.* (1994) *Nat. Genet.* 6(1): 84-89). Genome Systems clone #18720 was isolated in this screen and contains three repeats of human rDNA as assessed by restriction analysis. GS clone #18720, was digested with PmeI, a restriction enzyme unique to a single repeat of the human rDNA (45Kbp),

and then religated to form pPACrDNA (Figure 15). The insert in pPACrDNA was analyzed by restriction digests and sequence analysis of the 5' and 3' termini. The pPACrDNA, rDNA sequences are homologous to Genbank Accession #U13369, containing an insert of about 45 kB
5 comprising a single repeat beginning from the end of one repeat at ~33980 (relative to the Genbank sequence) through the beginning of the next repeat up to approximately 35120 (the repeat offset from that listed in the GenBank file). Thus, the rDNA sequence is just over 1 copy of the repeat extending from 33980 (+/-10bp) to the end of the first repeat
10 (43Kbp) and continuing into the second repeat to bp 35120 (+/-10bp).

B. TRANSFECTION AND *ACes* FORMATION.

Five hundred thousand MSU1.1 cells (Morgan et al., 1991, Exp. Cell Res., Nov;197(1):125-136; provided by Dr. Justin McCormick at Michigan State University) were plated per 6cm plate (3 plates total) and
15 allowed to grow overnight. The cells were 70-80% confluent the following day. One plate was transfected with 15µg pPACrDNA (linearized with *Pme* I) and 2µg pSV40attPsensePuro (linearized with *Sca* I; see Example 3). The remaining plates were controls and were transfected with either 20µg pBS (Stratagene) or 20µg
20 pSV40attBsensePuro (linearized with *Sca* I). All three plates were transfected using a CaPO₄ protocol.

C. SELECTION OF PUROMYCIN RESISTANT COLONIES

One day post-transfection the cells were "glycerol shocked" by the addition of PBS medium containing 10% glycerol for 30 seconds.
25 Subsequently, the glycerol was removed and replaced with fresh DMEM. Four days post-transfection selective medium was added. Selective medium contains 1µg/ml puromycin. The transfection plates were maintained at 37°C with 5% CO₂ in selective medium for 2 weeks at which point colonies could be seen on the plate transfected with

pPACrDNA and pSV40attPsensePuro. The colonies were ring-cloned from the plate on day 17 post-selection and expanded in selective medium for analysis. Only two colonies (M2-2d & M2-2b) were able to proliferate in the selective medium after cloning. No colonies were seen on the control plates after 37 days in selective medium.

D. ANALYSIS OF CLONES

FISH analysis was performed on the candidate clones to detect *ACes* formation. Metaphase spreads from the candidate clones were probed in multiple probe combinations. In one experiment, the probes used were biotin-labeled human alphoid DNA (pPACrDNA) and digoxigenin-labeled mouse major DNA (pFK161) as a negative control. Candidate M2-2d was single cell subcloned by flow sorting and the candidate subclones were reanalyzed by FISH. Subclone 1B1 of M2-2d was determined to be a platform *ACes* and is also designated human Platform ACE 0.1.

EXAMPLE 10

Site-specific integration of a marker gene onto a human platform ACE 0.1

The promoterless delivery method was used to deliver a promoterless blasticidin marker gene onto the human platform *ACes* with excellent results. The human *ACes* platform with a promoterless blasticidin marker gene resulted in 21 of 38 blasticidin resistant clones displaying a PCR product of the expected size from the population co-transfected with pLIT38attBSRpolyA10 and pCXLamIntR (Figure 8; SEQ ID NOs. 111 and 112). Whereas, the population transfected with pBlueScript resulted in 0 blasticidin resistant colonies.

A. CONSTRUCTION OF pLIT38attB-BSRpolyA10 & pLIT38attB-BSRpolyA2.

The vector pLITMUS 38 (New England Biolabs; U.S. Patent No. 5,691,140; SEQ ID NO: 119) was digested with *EcoRV* and ligated to

two annealed oligomers, which form an attB site (attB1 5'-TGAAGCCTGCTTTTTTATACTAACTTGAGCGAA-3' (SEQ ID NO:8); attB2 5'- TTCGCTCAAGTTAGTATAAAAAAGCAGGCTTCA-3'; SEQ ID NO:9).

This ligation reaction resulted in the vector pLIT38attB (SEQ ID NO: 120).

- 5 The blasticidin resistance gene and SV40 polyA site were PCR amplified with primers: 5BSD (ACCATGAAAACATTTAACATTTCTCAACA; SEQ ID NO:80) and SV40polyA (TTTATTTGTGAAATTTGTGATGCTATTGC; SEQ ID NO:81) using pPAC4 (Frengen, E., *et al.* (2000) Genomics 68 (2), 118-126; GenBank Accession No. U75992) as template. The blasticidin-
- 10 SV40polyA PCR product was then ligated into pLIT38attB at the *Bam*HI site, which was Klenow treated following digestion with *Bam*HI. pLIT38attB-BSDpolyA10 (SEQ ID NO: 111) and pLIT38attB-BSDpolyA2 (SEQ ID NO: 121) are the two resulting orientations of the PCR product ligated into the vector.

15 **B. TRANSFECTION OF MSU1.1 CELLS CONTAINING HUMAN PLATFORM ACE 0.1.**

- MSU1.1 cells containing human platform ACE 0.1 (see Example 9) were expanded and plated to five 10cm dishes with 1.3×10^6 cells per dish. The cells were incubated overnight in DMEM with 10% fetal bovine
- 20 serum, at 37°C and 5% CO₂. The following day the cells were transfected with 5µg of each plasmid as set forth in Table 3, for a total of 10µg of DNA per plate of cells transfected (see Table 3) using ExGen 500 *in vitro* transfection reagent (MBI fermentas, cat. no. R0511). The transfection was performed according to the manufacturers protocol.
- 25 Cells were incubated at 37°C with 5% CO₂ in DMEM with 10% fetal bovine serum following the transfection.

Table 3

Plate #	Plasmid 1	Plasmid 2	No. Bsd ^R Colonies
1	pBS	None	0
2	pCXLamInt	pLIT38attB-BSRpolyA10	16
3	pCXLamIntR	pLIT38attB-BSRpolyA10	40
4	pCXLamInt	pLIT38attB-BSRpolyA2	28
5	pCXLamIntR	pLIT38attB-BSRpolyA2	36

10 C. **SELECTION OF BLASTICIDIN RESISTANT CLONES.**

Three days following the transfection the cells were split from a 10 cm dish to two 15cm dishes. The cells were maintained in DMEM with 10% fetal bovine serum for 4 days in the 15 cm dishes. Seven days post-transfection blasticidin was introduced into the medium. Stably transfected cells were selected with 1 μ g/ml blasticidin. The number of colonies formed on each plate is listed in Table 3. These colonies were ring-cloned and expanded for PCR analysis. Upon expansion in blasticidin containing medium some clones failed to live and therefore do not have corresponding PCR data.

20 D. **PCR ANALYSIS**

Thirty-eight of the 40 clones from plate 3 grew after ring-cloning. Genomic DNA was isolated from these clones with the Promega Wizard Genomic cDNA purification kit, digested with *Eco*RI and used as template in a PCR reaction with the following primers: 3BSP – TTAATTTCTGGG TATATTTGAGTGGA (SEQ ID NO:82); 5PacSV40 – CTGTTAATTAAGTGTGGAA TGTGTGTCAGTTAGGGTG (SEQ ID NO:76). The PCR conditions were as follows. 100ng of genomic DNA was

amplified with 0.5 μ l Herculanase polymerase (Stratagene) in a 50 μ l reaction that contained 12.5pmole of each primer, 2.5mM of each dNTP, and 1X Herculanase buffer (Stratagene). The reactions were placed in a PerkinElmer thermocycler programmed as follows: Initial denaturation at 95°C for 10 minutes; 35 cycles of 94°C for 1 minute, 53°C for 1 minute, 72°C for 1 minute, and 72°C for 1 minute; Final extension for 10 minutes at 72°C; and 4°C hold. If pLIT38attB-BSRpolyA10 integrates onto the human platform ACE 0.1 correctly, PCR amplification with the above primers should yield an 804bp product. Twenty-one of the 38 clones from plate 10 3 produced a PCR product of the expected 804bp size.

EXAMPLE 11

Delivery of a Vector comprising a Promoterless Marker Gene and a gene encoding a therapeutic product to a Platform ACes

Platform ACes containing pSV40attPsensePURO (Figure 4) were constructed as set forth in Examples 3 and 4.

A. CONSTRUCTION OF DELIVERY VECTORS

1. Erythropoietin cDNA vector, p18EPOcDNA.

The erythropoietin cDNA was PCR amplified from a human cDNA library (E. Perkins *et al.*, 1999, *Proc. Natl. Acad. Sci. USA* 96(5): 2204-2209) using the following primers: EPO5XBA - TATCTAGAATGGGGGTGC ACGAATGTCCTGCC (SEQ ID NO: 83); EPO3BSI - TACGTACGTCATC TGTCCCCTGTCCTGCAGGC (SEQ ID NO: 84). The cDNA was amplified through two successive rounds of PCR using the following conditions: heat denaturation at 95°C for 3 minutes; 20 35 cycles of a 30 second denaturation (95°C), 30 seconds of annealing (60°C), and 1 minute extension (72°C); the last cycle is followed by a 7 minute extension at 72°C. BIO-X-ACT (BIOLINE) was used to amplify the erythropoietin cDNA from 2.5ng of the human cDNA library in the first round of amplification. Five μ l of the first amplification product was used 25

as template for the second round of amplification. Two PCR products were produced from the second amplification with Taq polymerase (Eppendorf), each product was cloned into pCR2.1-Topo (Invitrogen) and sequenced. The larger PCR product contained the expected cDNA
5 sequence for erythropoietin. The erythropoietin cDNA was moved from pTopoEPO into p18attBZeo(6XHS4)2eGFP (SEQ ID NO: 110). pTopoEPO was digested with BsiWI and XbaI to release a 588 bp EPO cDNA. BsrGI and BsiWI create compatible ends. The eGFP gene was removed from p18attBZeo(6XHS4)2eGFP by digestion with BsiWI and XbaI, the 8.3 Kbp
10 vector backbone was gel purified and ligated to the 588 bp EPO cDNA to create p18EPOcDNA (SEQ ID NO: 124).

2. Genomic erythropoietin vector, p18genEPO.

The erythropoietin genomic clone was PCR amplified from a human genomic library (Clontech) using the following primers: GENEPO3BSI -
15 CGTACGTCATCTGTCCCCT GTCCTGCA (SEQ ID NO: 85); GENEPO 5XBA -TCTAGAATGGGGGT GCACGGTGAGTACT (SEQ ID NO: 86). The reaction conditions for the amplification were as follows: heat denaturation for 3 minutes (95°C); 30 cycles of a 30 second denaturation (95°C), 30 seconds annealing (from 65°C decreasing 0.5°C per cycle to
20 50°C), and 3 minutes extension (72°C); 15 cycles of a 30 second denaturation (95°C), 30 seconds annealing (50°C), and 3 minute extension (72°C); the last cycle is followed by a 7 minute extension at 72°C. The erythropoietin genomic PCR product (2147 bp) was gel purified and cloned into pCR2.1Topo to create pTopogenEPO. Sequence
25 analysis revealed 2bp substitutions and insertions in the intronic sequences of the genomic clone of erythropoietin. A partial digest with XbaI and complete digest with BsiWI excised the erythropoietin genomic insert from pTopogenEPO. The resulting 2158 bp genomic erythropoietin fragment was ligated into the 8.3 Kbp fragment resulting from the

digestion of p18attBZeo(6XHS4)2eGFP (SEQ ID NO: 110) with XbaI and BsrGI to create p18genEPO (SEQ ID NO: 125).

B. TRANSFECTION AND SELECTION WITH DRUG

The erythropoietin genomic and cDNA genes were each moved
5 onto the platform ACes B19-38 (constructed as set forth in Example 3) by
co-transfecting with pCXLamIntR. Control transfections were also
performed using pCXLamInt (SEQ ID NO: 127) together with either
p18EPOcDNA (SEQ ID NO: 124) or p18genEPO (SEQ ID NO: 125).
Lipofectamine Plus was used to transfect the DNA's into B19-38 cells
10 according to the manufacturer's protocol. The cells were placed in
selective medium (DMEM with 10% FBS and Zeocin @ 500ug/ml) 48
hours post-transfection and maintained in selective medium for 13 days.
Clones were isolated 15 days post-transfection.

C. ANALYSIS OF CLONES (ELISA, PCR)

15 **1. ELISA Assays**

Thirty clones were tested for erythropoietin production by an ELISA
assay using a monoclonal anti-human erythropoietin antibody (R&D
Systems, Catalogue # MAB287), a polyclonal anti-human erythropoietin
antibody (R & D Systems, Catalogue # AB-286-NA) and alkaline
20 phosphatase conjugated goat-anti-rabbit IgG (heavy and light chains)
(Jackson ImmunoResearch Laboratories, Inc., Catalogue # 111-055-144).
The negative control was a Zeocin resistant clone isolated from B19-38
cells transfected with p18attBZeo(6XHS4) (SEQ ID NO: 117; no insert
control vector) and pCXLamIntR (SEQ ID NO: 112). The preliminary
25 ELISA assay was executed as follows: 1) Nunc-Immuno Plates (MaxiSorb
96-well, Catalogue # 439454) were coated with 75µl of a 1/200 dilution
(in Phosphate buffered Saline, pH 7.4 (PBS), Sigma Catalogue # P-3813)
of monoclonal anti-human erythropoietin antibody overnight at 4°C. 2)
The following day the plates were washed 3 times with 300µl PBS

containing 0.15% Tween 20 (Sigma, Catalogue # P-9416). 3) The plates were then blocked with 300 μ l of 1% Bovine Serum Albumin (BSA; Sigma Catalogue # A-7030) in PBS for 1 hour at 37°C. 4) Repeat the washes as in step 2. 5) The clonal supernatants (75 μ l per clone per well of 96-well plate) were then added to the plate and incubated for 1 hour at 37°C. The clonal supernatant analyzed in the ELISA assay had been maintained on the cells 7 days prior to analysis. 6) Repeat the washes of step 2. 7) Add 75 μ l of polyclonal anti-human erythropoietin antibody (1/250 dilution in dilution buffer (0.5% BSA, 0.01% Tween 20, 1X PBS, pH 7.4) and incubate 1 hour at 37°C. 8) Repeat washes of step 2. 9) Add 75 μ l of goat anti-rabbit conjugated alkaline phosphatase diluted 1/4000 in dilution buffer and incubate 1 hour at 37°C. 10) Repeat washes of step 2. 11) Add 75 μ l substrate, p-nitrophenyl phosphate (Sigma N2640), diluted to 1mg/ml in substrate buffer (0.1 Ethanolamine-HCl (Sigma, Catalogue # E-6133), 5mM MgCl₂ (Sigma, Catalogue # M-2393), pH 9.8). Incubate the plates in the dark for 1 hour at room temperature (22°C). 12) Read the absorption at 405nm (reference wavelength 495nm) on an Universal Microplate Reader (Bio-Tek Instruments, Inc., model # ELX800 UV). The erythropoietin standard curve was derived from readings of diluted human recombinant Erythropoietin (Roche, catalogue # 1-120-166; dilution range 125 - 7.8mUnits/ml). From this preliminary assay the 21 clones displaying the highest expression of erythropoietin were analyzed a second time in the same manner using medium supernatants that had been on the clones for 24 hours and a 1:3 dilution thereof.

25 **2. PCR Analysis**

Genomic DNA was isolated from the 21 clones with the best expression (as assessed by the initial ELISA assay above) as well as the B19-38 cell line and used for PCR analysis. Genomic DNA was isolated using the Wizard genomic DNA purification kit (Promega) according to the

manufacturers protocol. Amplification was performed on 100ng of genomic DNA as template with MasterTaq DNA Polymerase (Eppendorf) and the primer set 5PacSV40 – CTGTTAATTAAGTGTGGAATGTGTG TCAGTTAGGGTG (SEQ ID NO: 76) and Antisense Zeo -

- 5 TGAACAGGGTCACGTCGTCC (SEQ ID NO: 77). The amplification conditions were as follows: heat denaturation for 3 minutes (95°C); 30 cycles of a 30 second denaturation (95°C), 30 seconds annealing (from 65°C decreasing 0.5°C per cycle to 50°C), and 1 minutes extension (72°C); 15 cycles of a 30 second denaturation (95°C), 30 seconds
- 10 annealing (50°C), and 1 minute extension (72°C); the last cycle is followed by a 10 minute extension at 72°C. PCR products were size separated by gel electrophoresis. Of the 21 clones analyzed 19 produced a PCR product of 650 bp as expected for a site-specific integration event. All nineteen clones were the result of transformations with p19EPOcDNA
- 15 (5) or p18genEPO (14) and pCXLaIntR (i.e. mutant integrase). The remaining two clones, both of which were the result of transformation with p18genEPO (SEQ ID NO: 125) and pCXLaInt (i.e. wildtype integrase; SEQ ID NO: 127), produced a 400 bp PCR product.

EXAMPLE 12

20 Preparation of a Transformation Vector Useful for the Induction of Plant Artificial Chromosome Formation

- Plant artificial chromosomes (PACs) can be generated by introducing nucleic acid, such as DNA, which can include a targeting DNA, for example rDNA or lambda DNA, into a plant cell, allowing the cell
- 25 to grow, and then identifying from among the resulting cells those that include a chromosome with a structure that is distinct from that of any chromosome that existed in the cell prior to introduction of the nucleic acid. The structure of a PAC reflects amplification of chromosomal DNA, for example, segmented, repeat region-containing and heterochromatic

structures. It is also possible to select cells that contain structures that are precursors to PACs, for example, chromosomes containing more than one centromere and/or fragments thereof, and culture and/or manipulate them to ultimately generate a PAC within the cell.

5 In the method of generating PACs, the nucleic acid can be introduced into a variety of plant cells. The nucleic acid can include targeting DNA and/or a plant expressable DNA encoding one or multiple selectable markers (*e.g.*, DNA encoding bialophos (bar) resistance) or scorable markers (*e.g.*, DNA encoding GFP). Examples of targeting DNA
10 include, but are not limited to, *N. tabacum* rDNA intergenic spacer sequence (IGS) and *Arabidopsis* rDNA such as the 18S, 5.8S, 26S rDNA and/or the intergenic spacer sequence. The DNA can be introduced using a variety of methods, including, but not limited to *Agrobacterium*-mediated methods, PEG-mediated DNA uptake and electroporation using,
15 for example, standard procedures according to Hartmann *et al* [(1998) *Plant Molecular Biology* 36:741]. The cell into which such DNA is introduced can be grown under selective conditions and can initially be grown under non-selective conditions and then transferred to selective media. The cells or protoplasts can be placed on plates containing a
20 selection agent to grow, for example, individual calli. Resistant calli can be scored for scorable marker expression. Metaphase spreads of resistant cultures can be prepared, and the metaphase chromosomes examined by FISH analysis using specific probes in order to detect amplification of regions of the chromosomes. Cells that have artificial chromosomes with
25 functioning centromeres or artificial chromosomal intermediate structures, including, but not limited to, dicentric chromosomes, formerly dicentric chromosomes, minichromosomes, heterochromatin structures (*e.g.* sausage chromosomes), and stable self-replicating artificial chromosomal intermediates as described herein, are

identified and cultured. In particular, the cells containing self-replicating artificial chromosomes are identified.

The DNA introduced into a plant cell for the generation of PACs can be in any form, including in the form of a vector. An exemplary
5 vector for use in methods of generating PACs can be prepared as follows.

For the production of artificial chromosomes, plant transformation vectors, as exemplified by pAgIIa and pAgIIb, containing a selectable marker, a targeting sequence, and a scorable marker were constructed using procedures well known in the art to combine the various fragments.
10 The vectors can be prepared using vector pAg1 as a base vector and inserting the following DNA fragments into pAg1: DNA encoding β -glucuronidase under the control of the nopaline synthase (NOS) promoter fragment and flanked at the 3' end by the NOS terminator fragment, a fragment of mouse satellite DNA and an *N. tabacum* rDNA intergenic
15 spacer sequence (IGS). In constructing plant transformation vectors, vector pAg2 can also be used as the base vector.

1. Construction of pAg1

Vector pAg1 (SEQ. ID. NO: 89) is a derivative of the CAMBIA vector named pCambia 3300 (Center for the Application of Molecular
20 Biology to International Agriculture, i.e., CAMBIA, Canberra, Australia; www.cambia.org), which is a modified version of vector pCambia 1300 to which has been added DNA from the bar gene conferring resistance to phosphinothricin. The nucleotide sequence of pCambia 3300 is provided in SEQ. ID. NO: 90. pCambia 3300 also contains a lacZ alpha sequence
25 containing a polylinker region.

pAg1 was constructed by inserting two new functional DNA fragments into the polylinker of pCambia 3300: one sequence containing an attB site and a promoterless zeomycin resistance-encoding DNA flanked at the 3' end by a SV40 polyA signal sequence, and a second

sequence containing DNA from the hygromycin resistance gene (hygromycin phosphotransferase) conferring resistance to hygromycin for selection in plants. Although the zeomycin-SV40 polyA signal fusion is not expected to function in plant cells, it can be activated in mammalian
5 cells by insertion of a functional promoter element into the attB site by site-specific recombination catalyzed by the Lambda att integrase. Thus, the inclusion of the attB-zeomycin sequences allows for evaluation of functionality of plant artificial chromosomes in mammalian cells by activation of the zeomycin resistance-encoding DNA, and provides an att
10 site for further insertion of new DNA sequences into plant artificial chromosomes formed as a result of using pAg1 for plant transformation. The second functional DNA fragment allows for selection of plant cells with hygromycin. Thus, pAg1 contains DNA from the bar gene conferring resistance to phosphinothricin, DNA from the hygromycin resistance gene,
15 both resistance-encoding DNAs under the control of a separate cauliflower mosaic virus (CaMV) 35S promoter, and the attB-promoterless zeomycin resistance-encoding DNA.

pAg1 is a binary vector containing *Agrobacterium* right and left T-DNA border sequences for use in *Agrobacterium*-mediated transformation
20 of plant cells or protoplasts with the DNA located between the border sequences. pAg1 also contains the pBR322 Ori for replication in *E.coli*. pAg1 was constructed by ligating *Hind*III/*Pst*I-digested p3300attBZeo with *Hind*III/*Pst*I-digested pBSCaMV35SHyg as follows.

a. Generation of p3300attBZeo

25 Plasmid pCambia 3300 was digested with *Pst*I/*Eco*RI and ligated with *Pst*I/*Stu*I-digested pLITattBZeo (the nucleotide sequence of pLITattBZeo is provided in SEQ. ID. NO: 91), which contains DNA encoding the zeocin resistance gene and an attB Integrase recognition sequence, to generate p3300attBZeo which contains an attB site, a
30 promoterless

zeomycin resistance-encoding DNA flanked at the 3' end by a SV40 polyA signal, and a reconstructed *Pst*I site.

b. Generation of pBSCaMV35SHyg

A DNA fragment containing DNA encoding hygromycin phosphotransferase flanked by the CaMV 35S promoter and the CaMV 35S polyA signal sequence was obtained by PCR amplification of plasmid pCambia 1302 (GenBank Accession No. AF234298 and SEQ. ID. NO: 92). The primers used in the amplification reaction were as follows:
CaMV35SpolyA:
5'-CTGAATTAACGCCGAATTAATTCGGGGGATCTG-3' SEQ. ID. NO: 93
CaMV35Spr:
5'-CTAGAGCAGCTTGCCCAACATGGTGGAGCA-3' SEQ. ID. NO: 94
The 2100-bp PCR fragment was ligated with *Eco*RV-digested pBluescript II SK+ (Stratagene, La Jolla, CA, U.S.A.) to generate pBSCaMV35SHyg.

c. Generation of pAg1

To generate pAg1, pBSCaMV35SHyg was digested with *Hind*III/*Pst*I and ligated with *Hind*III/*Pst*I-digested p3300attBZeo. Thus, pAg1 contains the pCambia 3300 backbone with DNA conferring resistance to phosphinothricin and hygromycin under the control of separate CaMV 35S promoters, an attB-promoterless zeomycin resistance-encoding DNA recombination cassette and unique sites for adding additional markers, *e.g.*, DNA encoding GFP. The attB site can be used as described herein for the addition of new DNA sequences to plant artificial chromosomes, including PACs formed as a result of using the pAg1 vector, or derivatives thereof, in the production of PACs. The attB site provides a convenient site for recombinase-mediated insertion of DNAs containing a homologous att site.

2. pAG2

The vector pAg2 (SEQ. ID. NO: 95) is a derivative of vector pAg1 formed by adding DNA encoding a green fluorescent protein (GFP), under the control of a NOS promoter and flanked at the 3' end by a NOS polyA signal, to pAg1. pAg2 was constructed as follows. A DNA fragment
5 containing the NOS promoter was obtained by digestion of pGEM-T-NOS, or pGEMEasyNOS (SEQ. ID. NO: 96), containing the NOS promoter in the cloning vector pGEM-T-Easy (Promega Biotech, Madison, WI, U.S.A.), with *Xba*I/*Nco*I and was ligated to a *Xba*I/*Nco*I fragment of pCambia 1302 containing DNA encoding GFP (without the CaMV 35S promoter) to
10 generate p1302NOS (SEQ. ID. NO: 97) containing GFP-encoding DNA in operable association with the NOS promoter. Plasmid p1302NOS was digested with *Sma*I/*Bsi*WI to yield a fragment containing the NOS promoter and GFP-encoding DNA. The fragment was ligated with *Pme*I/*Bsi*WI-digested pAg1 to generate pAg2. Thus, pAg2 contains DNA
15 from the bar gene conferring resistance to phosphinothricin, DNA conferring resistance to hygromycin, both resistance-encoding DNAs under the control of a cauliflower mosaic virus 35S promoter, DNA encoding kanamycin resistance, a GFP gene under the control of a NOS promoter and the attB-zeomycin resistance-encoding DNA. One of skill in
20 the art will appreciate that other fragments can be used to generate the pAg1 and pAg2 derivatives and that other heterologous DNA can be incorporated into pAg1 and pAg2 derivatives using methods well known in the art.

3. pAgIIa and pAgIIb transformation vectors

25 Vectors pAgIIa and pAgIIb were constructed by inserting the following DNA fragments into pAg1: DNA encoding β -glucuronidase, the nopaline synthase terminator fragment, the nopaline synthase (NOS) promoter fragment, a fragment of mouse satellite DNA and an *N. tabacum*

rDNA intergenic spacer sequence (IGS). The construction of pAgIIa and pAgIIb was as follows.

An *N. tabacum* rDNA intergenic spacer (IGS) sequence (SEQ. ID. NO: 98; see also GenBank Accession No. YO8422; see also Borysyuk *et al.* (2000) *Nature Biotechnology* 18:1303-1306; Borysyuk *et al.* (1997) *Plant Mol. Biol.* 35:655-660; U.S. Patent Nos. 6,100,092 and 6,355,860) was obtained by PCR amplification of tobacco genomic DNA. The IGS can be used as a targeting sequence by virtue of its homology to tobacco rDNA genes; the sequence is also an amplification promoter sequence in plants. This fragment was amplified using standard PCR conditions (*e.g.*, as described by Promega Biotech, Madison, WI, U.S.A.) from tobacco genomic DNA using the primers shown below:

NTIGS-FI

5'- GTG CTA GCC AAT GTT TAA CAA GAT G- 3' (SEQ ID No. 99) and

15 NTIGS-RI

5'-ATG TCT TAA AAA AAA AAA CCC AAG TGA C- 3' (SEQ ID No. 100)

Following amplification, the fragment was cloned into pGEM-T Easy to give pIGS-I. A fragment of mouse satellite DNA (Msat1 fragment; GenBank Accession No. V00846; and SEQ ID No. 101) was amplified via PCR from pSAT-1 using the following primers:

MSAT-F1

5'- AAT ACC GCG GAA GCT TGA CCT GGA ATA TCG C -3'(SEQ ID No. 102) and

MSAT-RI

25 5'-ATA ACC GCG GAG TCC TTC AGT GTG CA T- 3' (SEQ ID No. 103)

This amplification added a *SacII* and a *HindIII* site at the 5' end and a *SacII* site at the 3' end of the PCR fragment. This fragment was then cloned into the *SacII* site in pIGS-1 to give pMIGS-1, providing a eukaryotic

centromere-specific DNA and a convenient DNA sequence for detection via FISH.

A functional marker gene containing a NOS-promoter:GUS:NOS terminator fusion was then constructed containing the NOS promoter
5 (GenBank Accession No. U09365; SEQ ID No. 104), *E. coli*
 β -glucuronidase coding sequence (from the GUS gene; GenBank
Accession No. S69414; and SEQ ID No. 105), and the nopaline synthase
terminator sequence (GenBank Accession No. U09365; SEQ ID No. 107).
The NOS promoter in pGEM-T-NOS was added to a promoterless GUS
10 gene in pBlueScript (Stratagene, La Jolla, CA, U.S.A.) using *NotI/Spel* to
form pNGN-1, which has the NOS promoter in the opposite orientation
relative to the GUS gene.

pMIGS-1 was digested with *NotI/Spel* to yield a fragment
containing the mouse major satellite DNA and the tobacco IGS which was
15 then added to *NotI*-digested pNGN-1 to yield pNGN-2. The NOS promoter
was then re-oriented to provide a functional GUS gene, yielding pNGN-3,
by digestion and religation with *Spel*. Plasmid pNGN-3 was then digested
with *HindIII*, and the *HindIII* fragment containing the β -glucuronidase
coding sequence and the rDNA intergenic spacer, along with the Msat
20 sequence, was added to pAg-1 to form pAgIIa (SEQ ID NO: 108), using
the unique *HindIII* site in pAg1 located near the right T-DNA border of
pAg1, within the T-DNA region.

Another plasmid vector, referred to as pAgIIb, was also recovered,
which contained the inserted *HindIII* fragment (SEQ ID NO: 108) in the
25 opposite orientation relative to that observed in pAgIIa. Thus, pAgIIa and
pAgIIb differ only in the orientation of the *HindIII* fragment containing the
mouse major satellite sequence, the GUS DNA sequence and the IGS
sequence. The nucleotide sequence of pAgIIa is provided in SEQ. ID. NO:
109.

-138-

Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

WHAT IS CLAIMED IS:

1. A eukaryotic chromosome comprising one or a plurality of *att* site(s), wherein:
 - an *att* site is heterologous to the chromosome; and
 - 5 an *att* site permits site-directed integration in the presence of lambda integrase.
2. The eukaryotic chromosome of claim 1, wherein the *att* sites are selected from the group consisting of *attP* and *attB* or *attL* and *attR*, or variants thereof.
- 10 3. The eukaryotic chromosome of claim 1 that is an artificial chromosome.
4. The eukaryotic chromosome of claim 1 that is an artificial chromosome expression system (ACes).
5. The eukaryotic chromosome of claim 4 that is predominantly
15 heterochromatin.
6. The chromosome of claim 1 that is an artificial chromosome that contains no more than about 30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% euchromatin.
7. The chromosome of claim 1 that is a plant chromosome.
- 20 8. The chromosome of claim 1 that is an animal chromosome.
9. The chromosome of claim 7 that is a plant artificial chromosome.
10. The chromosome of claim 8 that is an animal artificial chromosome.
- 25 11. The chromosome of claim 8 that is a mammalian chromosome.
12. The chromosome of claim 11 that is a mammalian artificial chromosome.

13. The chromosome of claim 6 that is an artificial chromosome expression system (*ACes*).

14. A platform artificial chromosome expression system (*ACes*) comprising one or a plurality of sites that participate in recombinase
5 catalyzed recombination.

15. The *ACes* of claim 14 that contains one site.

16. The *ACes* of claim 14 that is predominantly heterochromatin.

17. The *ACes* of claim 14 that contains no more than about
30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% euchromatin.

10 18. The *ACes* of claim 14 that is a plant *ACes*.

19. The *ACes* of claim 14 that is an animal *ACes*.

20. The *ACes* of claim 14 that is selected from a fish, insect, reptile, amphibian, arachnid or a mammalian *ACes*.

21. The *ACes* of claim 14 that is a fish *ACes*.

15 22. The artificial chromosome expression system (*ACes*) of claim 14, wherein the recombinase and site(s) are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase system of phage Mu, the Cin recombinase system, the Pin recombinase system of *E. coli* and the R/RS
20 system of the pSR1 plasmid, or any combination thereof.

23. A method of introducing heterologous nucleic acid into a chromosome, comprising:

contacting a chromosome of any of claims 1 or 14 with a nucleic acid molecule comprising both the heterologous nucleic acid and a
25 recombination site, in the presence of a recombinase that promotes recombination between the sites in the chromosome and in the nucleic acid molecule.

24. The method of claim 23, wherein the recombinase is selected from the group consisting of Cre, Gin, Cin, Pin, FLP, a phage integrase and R from the pSR1 plasmid.

5 25. The method of claim 23, wherein the nucleic acid molecule encodes a therapeutic protein, antisense nucleic acid, or comprises an artificial chromosome.

26. The method of claim 25, wherein the nucleic acid molecule comprises a yeast artificial chromosomes (YAC), a bacterial artificial chromosome (BAC) or an insect artificial chromosome (IAC).

10 27. A combination, comprising, the chromosome of claim 1 and a first vector comprising the cognate recombination site, wherein the cognate recombination site is a site that recombines with the site engineered into the chromosome.

15 28. The combination of claim 27, further comprising nucleic acid encoding a recombinase, wherein the nucleic acid is on a second vector or on the first vector, or on the ACes under an inducible promoter.

29. The combination of claim 28, wherein the recombinase and sites are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase system of phage Mu, the Pin recombinase system of *E. coli* and the R/RS system of the pSR1 plasmid, or any combination thereof.

30. The combination of claim 28, wherein a vector is the plasmid pCXLamIntR.

25 31. The combination of claim 27, wherein a vector is the plasmid pDsRedN1-attB.

32. A kit, comprising the combination of claim 27 and optionally instructions for introducing heterologous nucleic acid into the chromosome.

33. A method for introducing heterologous nucleic acid into a platform artificial chromosome, comprising:

(a) mixing an artificial chromosome comprising at least a first recombination site and a vector comprising at least a second recombination site and the heterologous nucleic acid;

(b) incubating the resulting mixture in the presence of at least one recombination protein under conditions whereby recombination between the first and second recombination sites is effected, thereby introducing the heterologous nucleic acid into the artificial chromosome.

34. The method of claim 33, wherein the artificial chromosome is an ACes.

35. The method of claim 33, wherein said mixing step (a) is conducted in cells ex vivo.

36. The method of claim 33, wherein said mixing step (a) is conducted extracellularly in an in vitro reaction mixture.

37. The method of claim 33, wherein the at least one recombination protein is encoded by a bacteriophage selected from the group consisting of bacteriophage lambda, phi 80, P22, P2, 186, P4 and P1.

38. The method of claim 37, wherein the at least one recombination protein is encoded by bacteriophage lambda, or mutants thereof.

39. The method of claim 33, wherein at least one recombination protein is selected from the group consisting of Int, IHF, Xis and Cre, $\gamma\delta$, Tn3 resolvase, Hin, Gin, Cin and Flp.

40. The method of claim 32, wherein the recombination sites are selected from the group consisting of att and lox P sites.

41. The method of claim 33, wherein the first and/or second recombination site contains at least one mutation that removes one or more stop codons.

42. The method of claim 33, wherein the first and/or second recombination site contains at least one mutation that avoids hairpin formation.

43. The method of claim 33, wherein the first and/or second recombination site comprises at least a first nucleic acid sequence selected from the group consisting of SEQ ID NOs:41-56:

- 10 a) RKYCWGCTTTYKTRTACNAASTSGB (m-att) (SEQ ID NO:41);
- b) AGCCWGCTTTYKTRTACNAACTSGB (m-attB) (SEQ ID NO:42);
- c) GTTCAGCTTCKTRTACNAACTSGB (m-attR) (SEQ ID NO:43);
- d) AGCCWGCTTCKTRTACNAAGTSGB (m-attL) (SEQ ID NO:44);
- e) GTTCAGCTTTYKTRTACNAAGTSGB (m-attP1) (SEQ ID NO:45);
- 15 f) AGCCTGCTTTTTGTACAAACTTGT (attB1) (SEQ ID NO:46);
- g) AGCCTGCTTCTTGTACAAACTTGT (attB2) (SEQ ID NO:47);
- h) ACCCAGCTTCTTGTACAAACTTGT (attB3) (SEQ ID NO:48);
- i) GTTCAGCTTTTTGTACAAACTTGT (attR1) (SEQ ID NO:49);
- j) GTTCAGCTTCTTGTACAAACTTGT (attR2) (SEQ ID NO:50);
- 20 k) GTTCAGCTTCTTGTACAAAGTTGG (attR3) (SEQ ID NO:51);
- l) AGCCTGCTTTTTGTACAAAGTTGG (attL1) (SEQ ID NO:52);
- m) AGCCTGCTTCTTGTACAAAGTTGG (attL2) (SEQ ID NO:53);
- n) ACCCAGCTTCTTGTACAAAGTTGG (attL3) (SEQ ID NO:54);
- o) GTTCAGCTTTTTGTACAAAGTTGG (attP1) (SEQ ID NO:55);
- 25 p) GTTCAGCTTCTTGTACAAAGTTGG (attP2, P3) (SEQ ID NO: 56);

and a corresponding or complementary DNA or RNA sequence, wherein R=A or G, K=G or T/U, Y=C or T/U, W=A or T/U, N=A or C or G or T/U, S=C or G, and B=C or G or T/U; and

the core region does not contain a stop codon in one or more reading frames.

44. The method of claim 33, wherein the first and/or second recombination site comprises at least a first nucleic acid sequence
5 selected from the group consisting of a mutated att recombination site containing at least one mutation that enhances recombinational specificity, a complementary DNA sequence thereto, and an RNA sequence corresponding thereto.

45. The method of claim 33, wherein the vector comprising the
10 second site further encodes at least one selectable marker.

46. The method of claim 45, wherein the marker is a promoterless marker, which, upon recombination is under the control of a promoter and is thereby expressed.

47. The method of claim 46, wherein the first recombination site
15 is attP and is in the sense orientation prior to recombination.

48. The method of claim 46, wherein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic, fluorescent, or capable of being bound by an antibody and FACs sorted.

20 49. The method of claim 48, wherein the selectable marker is selected from the group consisting of green fluorescent protein (GFP), red fluorescent protein (RFP), blue fluorescent protein (BFP), and *E. coli* histidinol dehydrogenase (hisD).

50. A cell comprising, the chromosome of claim 1.

25 51. The cell of claim 50, wherein the cell is a nuclear donor cell.

52. The cell of claim 50, wherein the cell is a stem cell.

53. The stem cell of claim 52, wherein said stem cell is human and is selected from the group consisting of a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell and an embryonic stem cell.

54. The cell of claim 50, wherein the cell is mammalian.

55. The cell of claim 54, wherein the mammal is selected from the group consisting of humans, primates, cattle, pigs, rabbits, goats, sheep, mice, rats, guinea pigs, hamsters, cats, dogs, and horses.

5 56. The cell of claim 50, wherein the cell is a plant cell.

57. A cell comprising the platform *ACes* of claim 14.

58. The cell of claim 57, wherein the cell is a nuclear donor cell.

59. The cell of claim 57, wherein the cell is a stem cell.

60. The stem cell of claim 59, wherein said stem cell is human
10 and is selected from the group consisting of a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell and an embryonic stem cell.

61. A human mesenchymal cell comprising an artificial chromosome.

62. The human mesenchymal cell of claim 61, wherein said
15 artificial chromosome is an *ACes*.

63. The human mesenchymal cell of claim 62, wherein the *ACes* is a platform-*ACes*.

64. A method for introducing heterologous nucleic acid into the mesenchymal cell of claim 63, comprising:

20 (a) introducing into the cell of claim 63, wherein the platform-*ACes* has a first recombination site, a vector comprising at least a second recombination site and the heterologous nucleic acid;

(b) incubating the resulting mixture in the presence of at least one recombination protein under conditions whereby recombination between
25 the first and second recombination sites is effected, thereby introducing the heterologous nucleic acid into the platform-*ACes* within the mesenchymal cell.

65. A lambda-intR mutein comprising a glutamic acid to arginine change at position 174 of wild-type lambda-intR.

66. The lambda-intR mutein of claim 65, wherein the lambda-intR mutein comprises SEQ ID NO:37.

67. The method of claim 46 wherein the promoterless marker is transcriptionally downstream of the heterologous nucleic acid, wherein
5 the heterologous nucleic acid encodes a heterologous protein, and wherein the expression level of the selectable marker is transcriptionally linked to the expression level of the heterologous protein.

68. The method of claim 67, wherein the selectable marker and the heterologous nucleic acid are transcriptionally linked by the presence
10 of a IRES between them.

69. The method of claim 68, wherein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic or fluorescent.

70. The method of claim 69, wherein the selectable marker is selected from the group consisting of green fluorescent protein (GFP), red fluorescent protein (RFP), blue fluorescent protein (BFP), and *E. coli* histidinol dehydrogenase.
15

71. The method of claim 67 further comprising expressing the
20 heterologous protein and isolating the heterologous protein.

72. A method for producing a transgenic animal, comprising introducing a platform-ACes into an embryonic cell.

73. The method of claim 72, wherein the embryonic cell is a stem cell.

74. The method of claim 72, wherein the embryonic cell is in an embryo.
25

75. The method of claim 72, wherein the platform-ACes comprises heterologous nucleic acid that encodes a therapeutic product.

76. The method of claim 72, wherein the transgenic animal is a fish, insect, reptile, amphibians, arachnid or mammal.

77. The method of claim 72, wherein the *ACes* is introduced by cell fusion, lipid-mediated transfection by a carrier system, microinjection, microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer.

78. A transgenic animal produced by the method of claim 72.

79. A cell line useful for making a library of *ACes*, comprising a multiplicity of heterologous recombination sites randomly integrated throughout the endogenous chromosomes.

80. A method of making a library of *ACes* comprising random portions of a genome, comprising introducing one or more *ACes* into the cell line of claim 79, under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites within the cell's chromosomal DNA; and isolating said multiplicity of *ACes*, thereby producing a library of *ACes* whereby multiple *ACes* have different portions of the genome within.

81. A library of cells useful for genomic screening, said library comprising a multiplicity of cells, wherein each cell comprises an *ACes* having a mutually exclusive portion of a chromosomal nucleic acid therein.

82. The library of cells of claim 81, wherein the cells of the library are from a different species than the chromosomal nucleic acid within the *ACes*.

83. A method of making one or more cell lines, comprising
a) integrating into endogenous chromosomal DNA of a selected cell species, a multiplicity of heterologous recombination sites,

b) introducing a multiplicity of *ACes* under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites integrated within the cell's endogenous chromosomal DNA;

5 c) isolating said multiplicity of *ACes*, thereby producing a library of *ACes* whereby a multiplicity of *ACes* have mutually exclusive portions of the endogenous chromosomal DNA therein;

 d) introducing the isolated multiplicity of *ACes* of step c) into a multiplicity of cells, thereby creating a library of cells;

10 e) selecting different cells having mutually exclusive *ACes* therein and clonally expanding or differentiating said different cells into clonal cell cultures, thereby creating one or more cell lines.

 84. The method of claim 23, wherein the nucleic acid molecule with a recombination site is a PCR product.

15 85. Method of claim 23 wherein the recombinase is a protein and the recombination event occurs in vitro.

 86. The method of claim 33, wherein the vector is a PCR product comprising a second recombination site.

20 87. The lambda-intR mutein of claim 65, wherein the mutein further comprises an amino acid signal for nuclear localization.

 88. The lambda-intR mutein of claim 65, wherein the mutein further comprises an epitope tag for protein purification.

 89. A modified iron-induced promoter comprising SEQ ID NO:128.

25 90. A plasmid or expression cassette comprising the promoter of claim 89.

 91. A vector, comprising:
 a recognition site for recombination; and

a sequence of nucleotides that targets the vector to an amplifiable region of a chromosome.

92. The vector of claim 91, wherein the amplifiable region comprises heterochromatic nucleic acid.

5 93. The vector of claim 91, wherein the amplifiable region comprises rDNA.

94. The vector of claim 93, wherein the rDNA comprises an intergenic spacer.

95. The vector of claim 91, further comprising nucleic acid
10 encoding a selectable marker that is not operably associated with any promoter.

96. The vector of claim 91, wherein the chromosome is a mammalian chromosome.

97. The vector of claim 91, wherein the chromosome is a plant
15 chromosome.

98. A cell of claim 57 that is a plant cell, wherein the ACes platform is a MAC.

99. The plant cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

20 100. The plant cell of claim 99, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

101. A cell of claim 57 that is an animal cell, wherein the ACes platform is a plant artificial chromosome (PAC).

25 102. The cell of claim 101 that is a mammalian cell.

103. The cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

104. The cell of claim 102, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

105. The cell of claim 104, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

106. A method, comprising:
5 introducing a vector of claim 91 into a cell;
 growing the cells; and
 selecting a cell comprising an artificial chromosome that comprises one or more repeat regions.

107. The method of claim 106, wherein sufficient portion of the
10 vector integrates into a chromosome in the cell to result in amplification of chromosomal DNA.

108. The method of claim 106, wherein the artificial chromosome is an *ACes*.

109. A method for screening, comprising:
15 contacting a cell comprising a reporter *ACes* with test compounds or known compounds, wherein:
 the reporter *ACes* comprises one or a plurality of reporter constructs;
 a reporter construct comprises a reporter gene in operative linkage
20 with a regulatory region responsive to test or known compounds; and
 detecting any increase or decrease in signal output from the reporter, wherein a change in the signal is indicative of activity of the test or known compound on the regulatory region.

110. The method of claim 109, wherein the reporter is operatively
25 linked to a promoter that controls expression of a gene in a signal transduction pathway, whereby activation or reduction in the signal indicates that the pathway is activated or down-regulated by the test compound.

111. The method of claim 109, wherein the reporter in the construct encodes drug resistance or encodes a fluorescent protein.

112. The method of claim 111, wherein the fluorescent protein is selected from the group consisting of red, green and blue fluorescent
5 proteins.

113. The method of claim 109, wherein the ACes comprises a plurality of reporter-linked constructs, each with a different reporter, whereby the pathway(s) affected by the test compounds can be elucidated.

10 114. The method of claim 109, wherein a reporter is operatively linked to a promoter that is transcriptionally regulated in response to DNA damage, and the test compounds are genotoxicants.

115. The method of claim 114, wherein the DNA damage is induced by apoptosis, necrosis or cell-cycle perturbations.

15 116. The method of claim 114, wherein unknown compounds are screened to assess whether they are genotoxicants.

117. The method of claim 114, wherein the promoter is a cytochrome P450-profiled promoter.

20 118. The method of claim 114, wherein the cell is in a transgenic animal and toxicity is assessed in the animal.

119. The method of claim 109, wherein:

the cell is a patient cell sample; the patient has a disease;

the regulatory region is one targeted by a drug or drug regimen;

and

25 the method assesses the effectiveness of a treatment for the disease for the particular patient.

120. The method of claim 119, wherein the cell is a tumor cell.

121. The method of claim 109, wherein the cell is a stem cell or a progenitor cell, whereby expression of the reporter is operatively linked to

a regulatory region expressed in the cells to thereby identify stem cells or progenitor cell.

122. The method of claim 109, wherein the cell is in an animal;
and the method comprises whole-body imaging to monitor expression of
5 the reporter in the animal.

123. A reporter ACes comprises one or a plurality of reporter constructs, wherein the reporter construct comprises a reporter gene in operative linkage with a regulatory region responsive to test or known compounds.

Generation of ACes for Platform
Chromosome Engineering

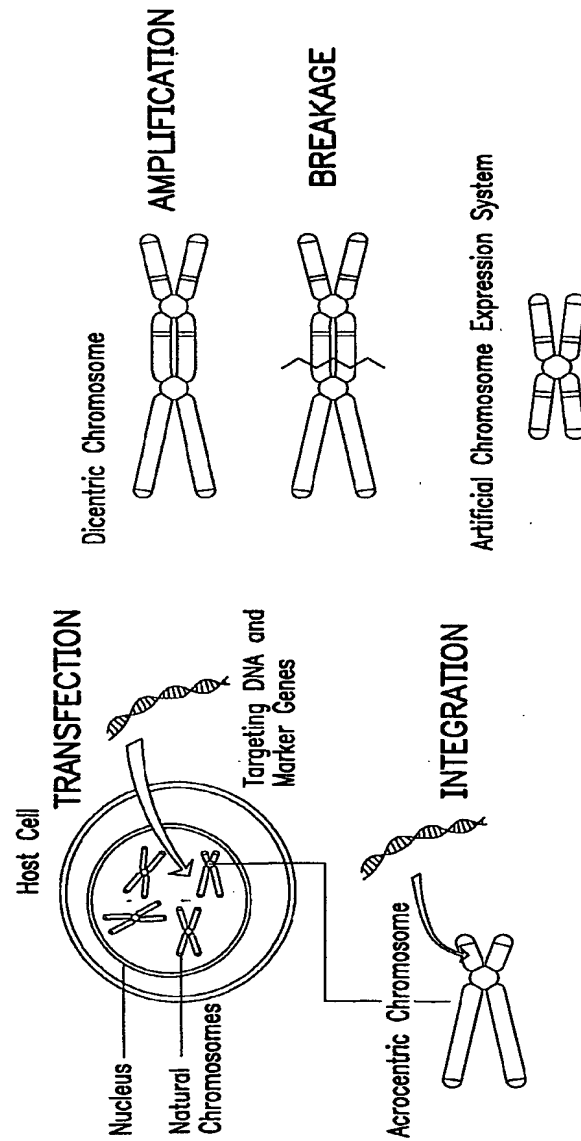


FIG. 1

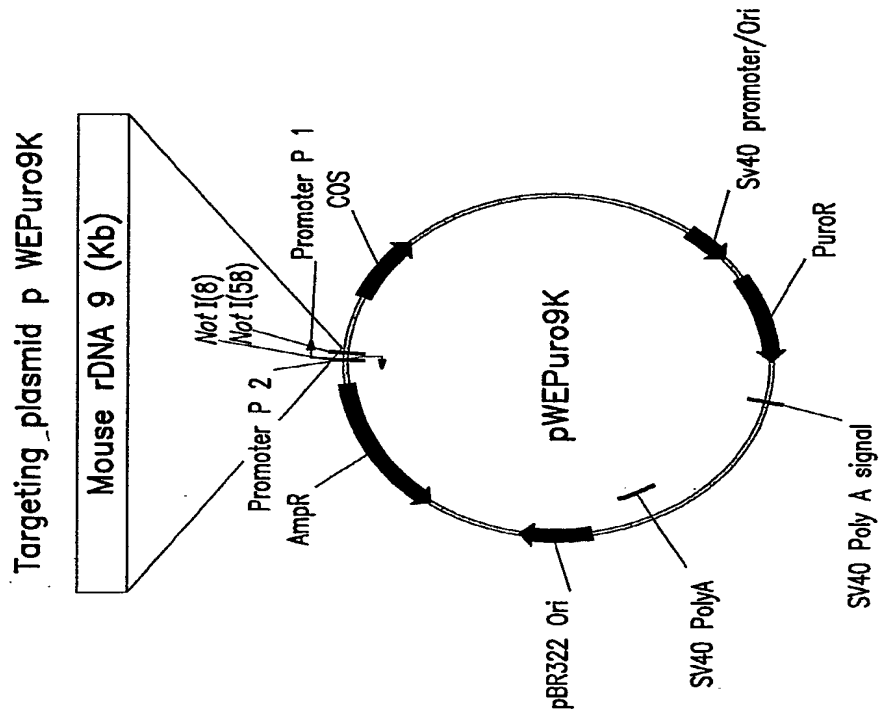


FIG. 2

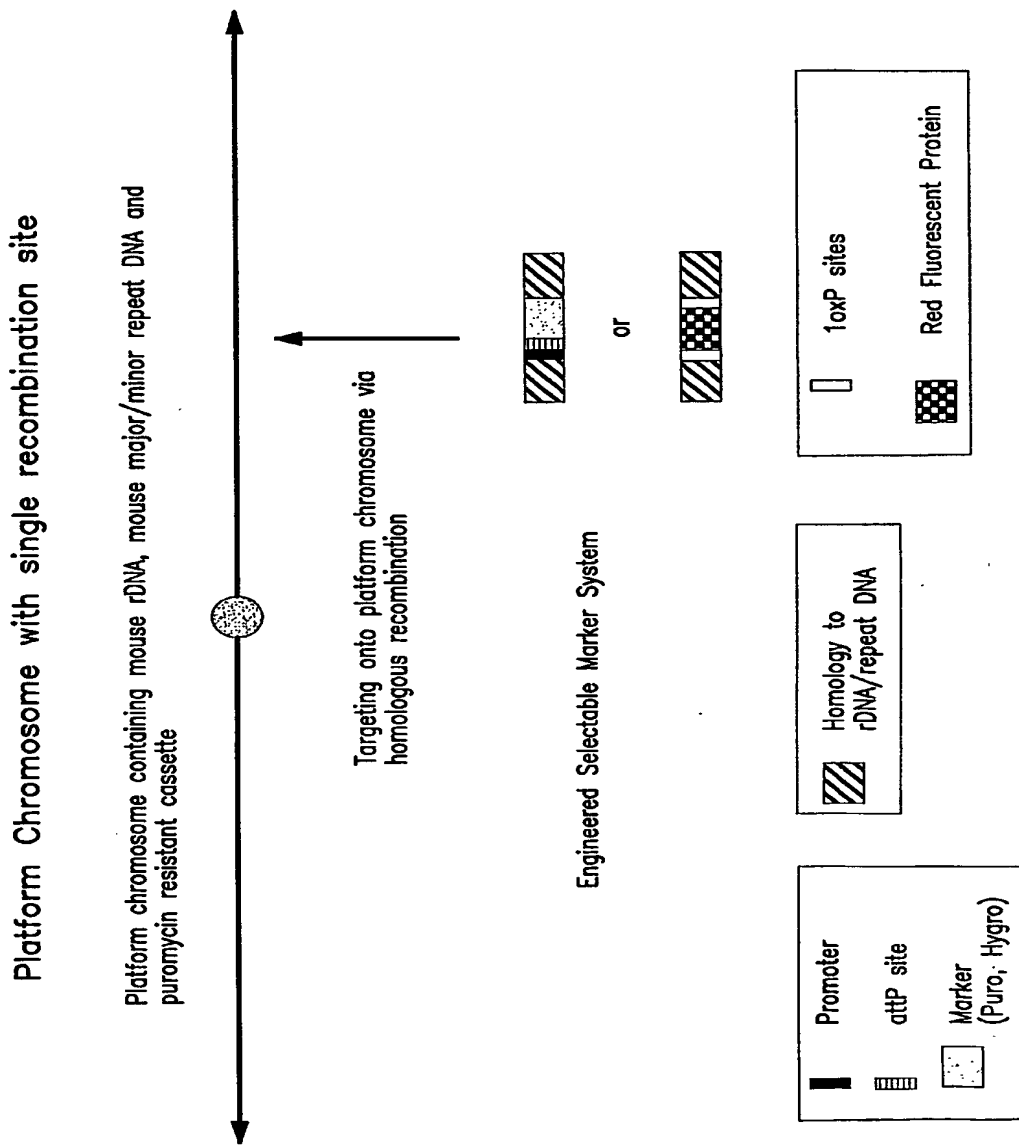


FIG. 3

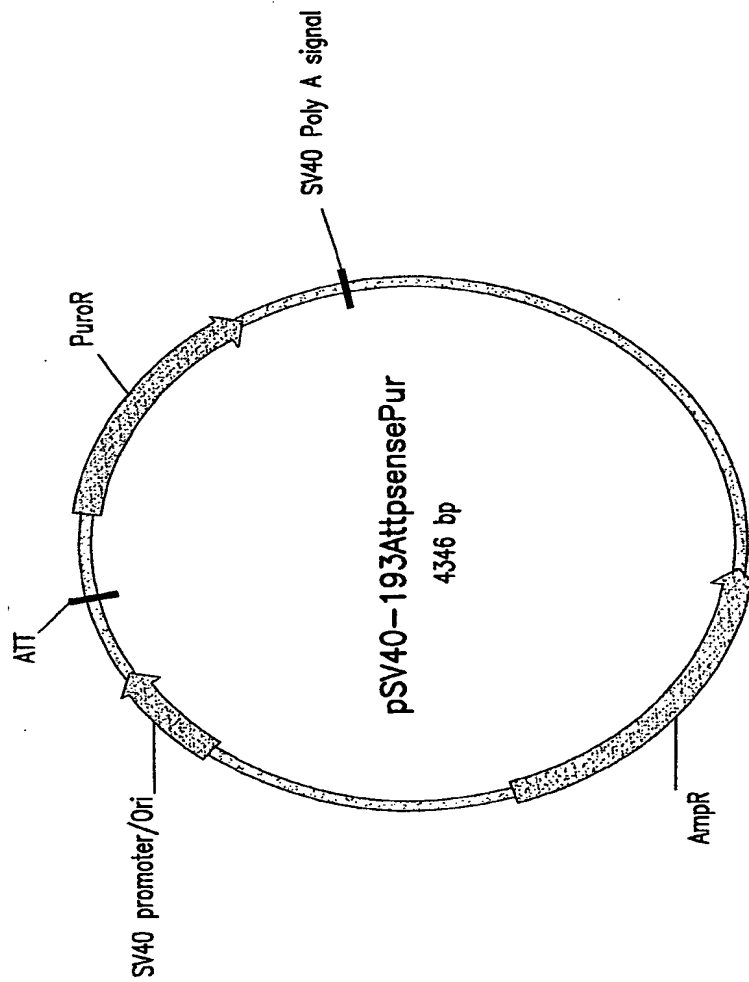


FIG. 4

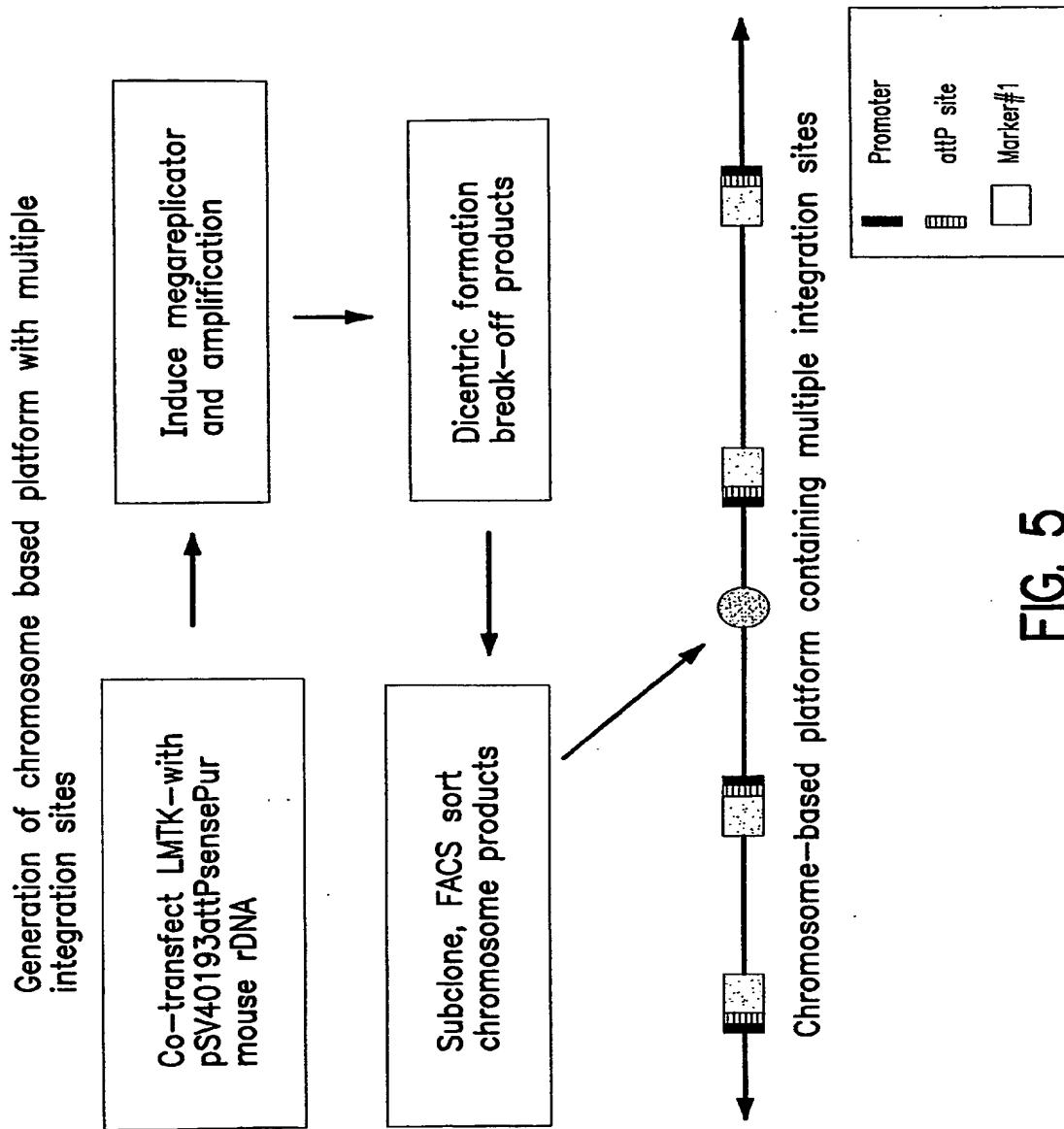


FIG. 5

λ integrase recombination

$\text{attP} \times \text{attB} \rightleftharpoons \text{attL} \times \text{attR}$

Core Region	
attP	CAGCTTTT <u>T</u> TATACTAAGTTG
attB	CTGCTTTT <u>T</u> TATACTAAGTTG
attL	CTGCTTTT <u>T</u> TATACTAAGTTG
attR	CAGCTTTT <u>T</u> TATACTAAGTTG

FIG. 6

λ INT recombination on artificial chromosome

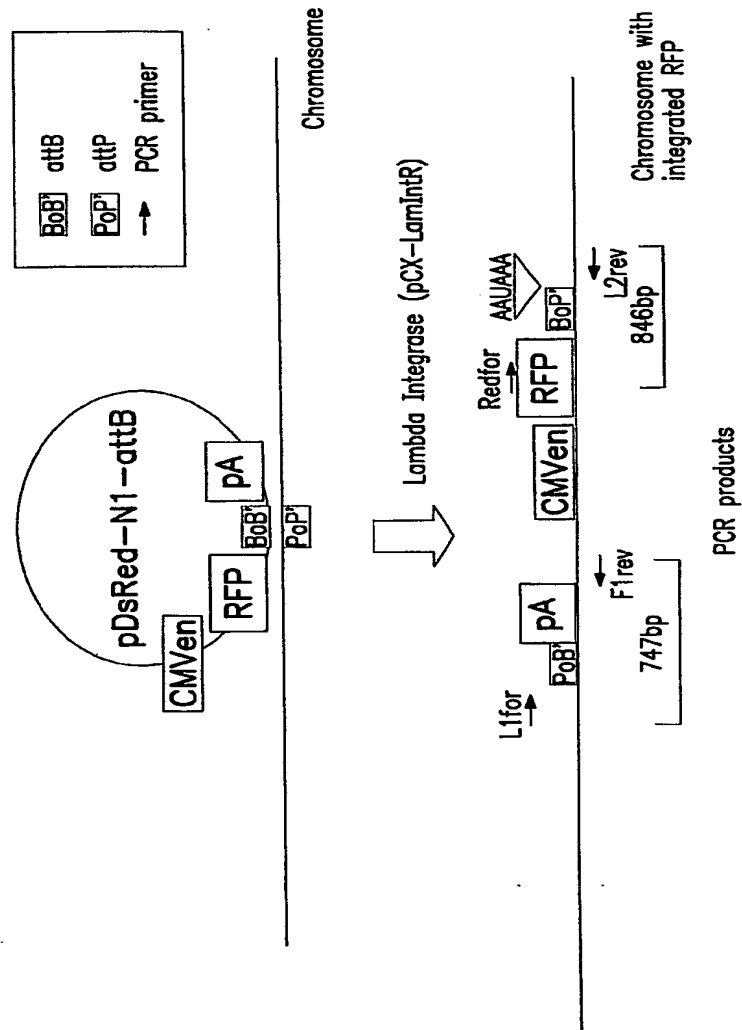


FIG. 7

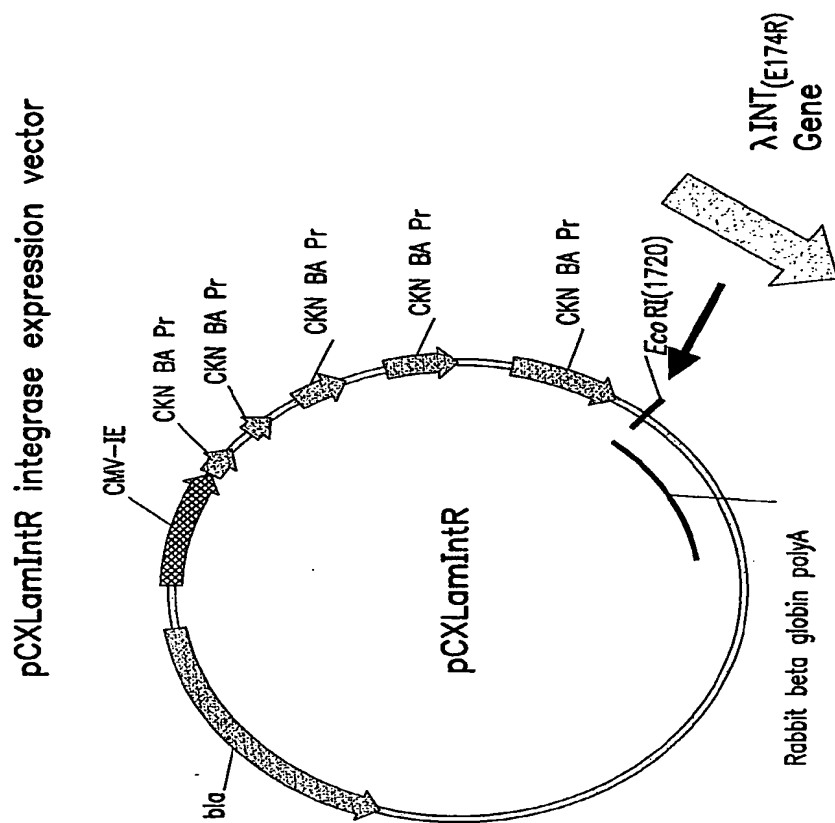


FIG. 8



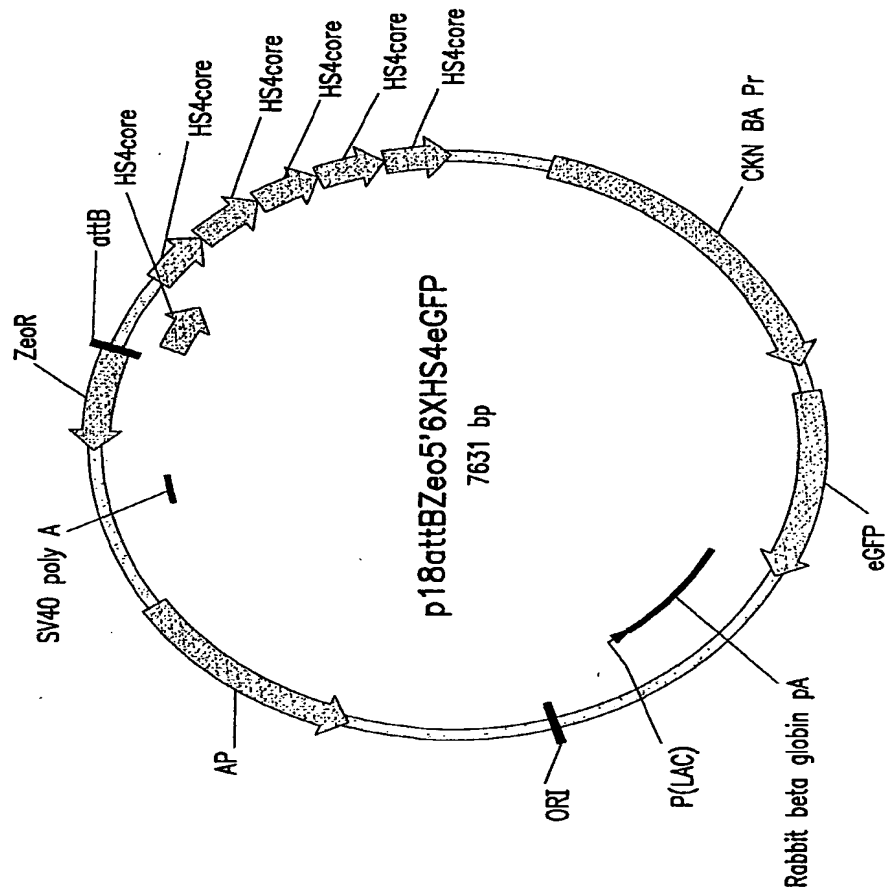


FIG. 10

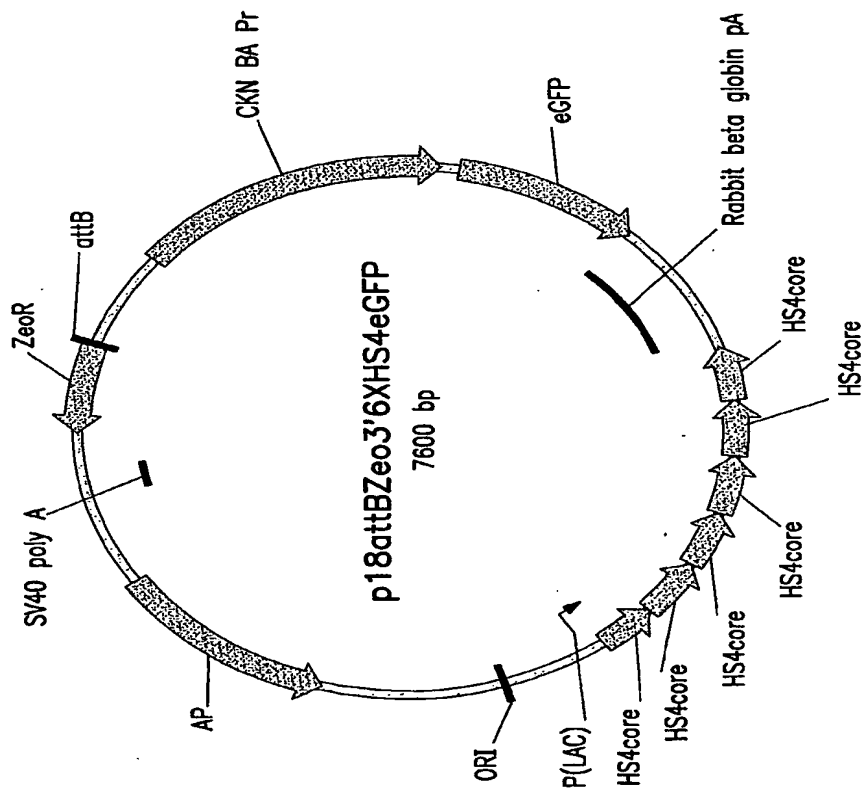


FIG. 11

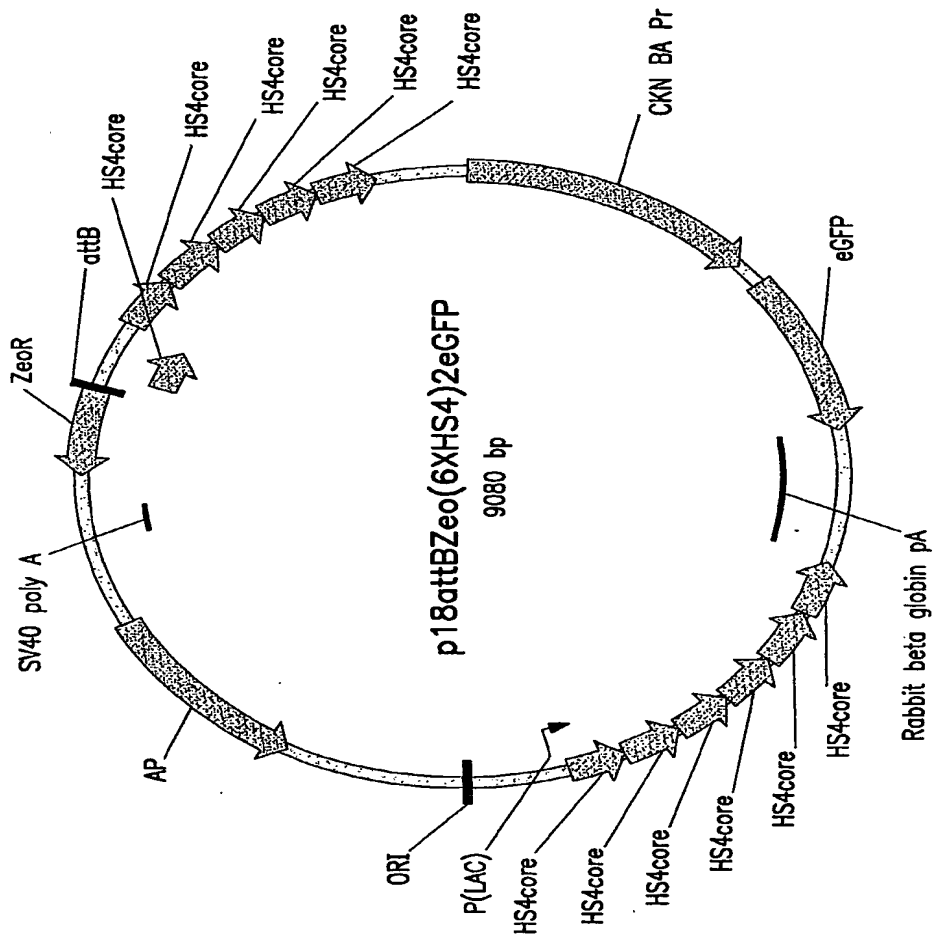


FIG. 12

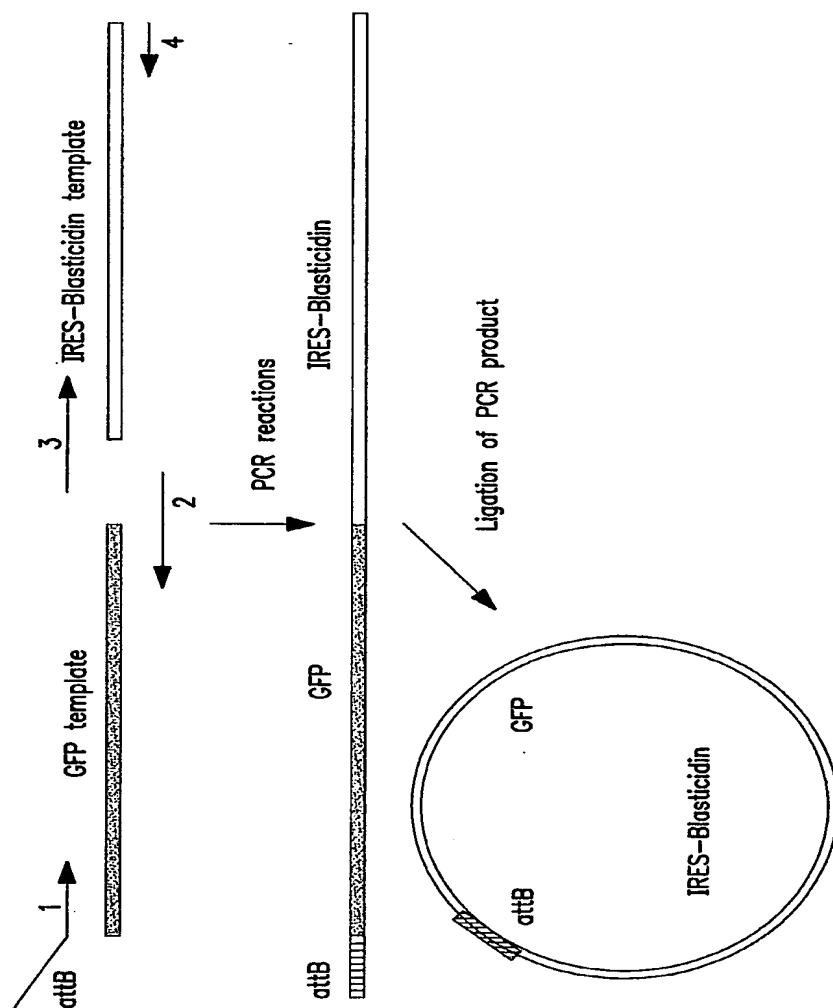


FIG. 13

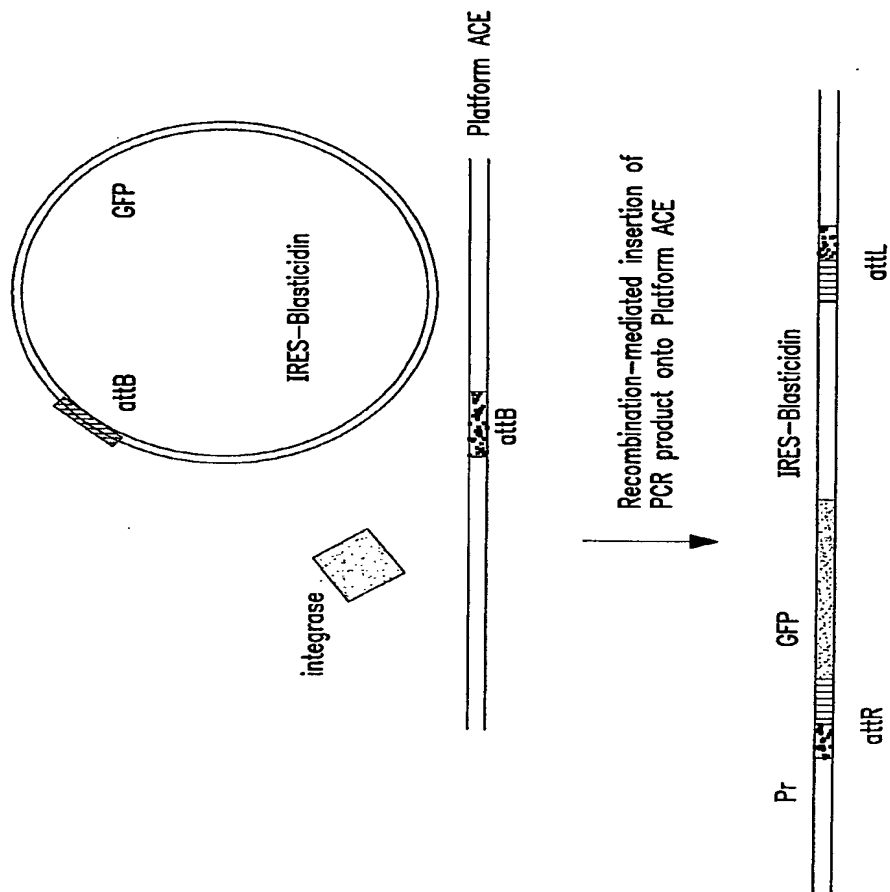


FIG. 14

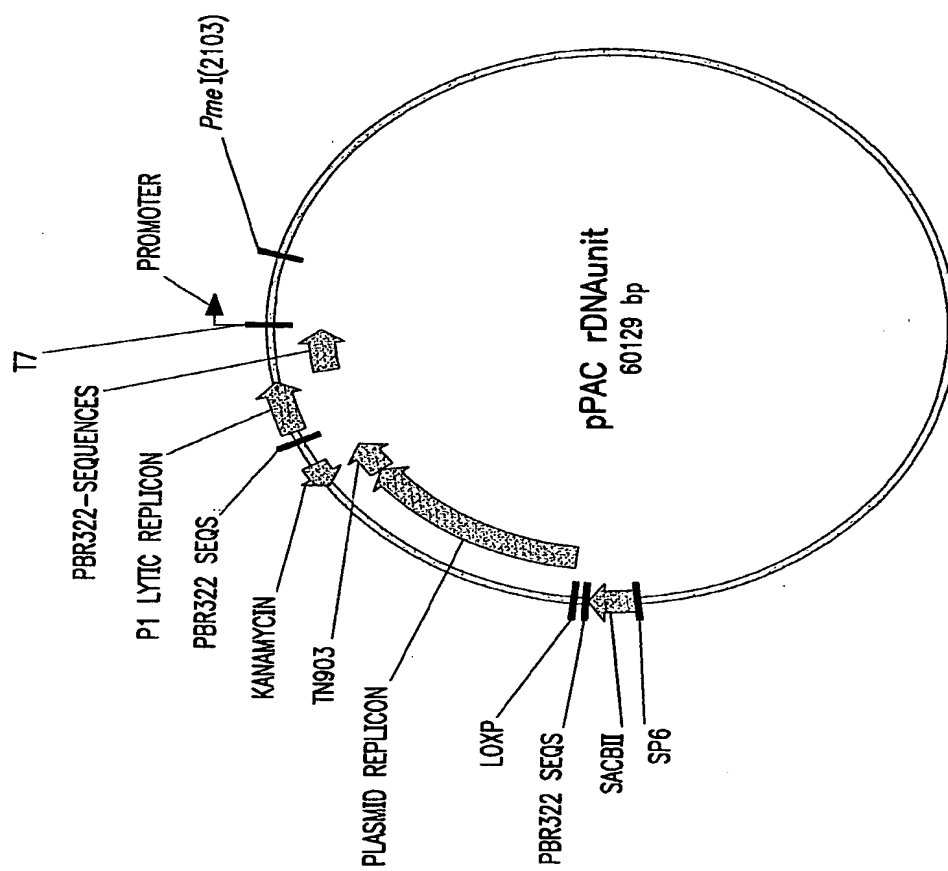


FIG. 15

SEQUENCE LISTING

<110> CHROMOS MOLECULAR SYSTEMS, INC.

Perkins, Edward
Perez, Carl
Lindenbaum, Michael
Greene, Amy
Leung, Josephine
Fleming, Elena
Stewart, Sandra
Shellard, Joan

<120> CHROMOSOME-BASED PLATFORMS

<130> 24601-420PC

<140> Not Yet Assigned

<141> Herewith

<150> 60/294,758

<151> 2001-05-30

<150> 60/366,891

<151> 2002-03-21

<160> 129

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer: attPUP

<400> 1

ccttgcgcta atgctctggt acagg

25

<210> 2

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer: attPDWN

<400> 2

cagaggcagg gagtgggaca aaattg

26

<210> 3

<211> 35

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer: Lamint 1

<400> 3

ttcgaattca tgggaagaag gcgaagtcac gaggc

35

<210> 4

<211> 34

<212> DNA

<213> Artificial Sequence

<220>
<223> Primer: Lamint 2

<400> 4
ttcgaattct tatttgattt caattttgtc ccac 34

<210> 5
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 5
cggacaatgc ggttgtgcgt 20

<210> 6
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 6
cgcgcagcaa aatctagagt aaggagatca agacttacgg ctgacg 46

<210> 7
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> LambdaINTER174rev

<400> 7
cgtcagccgt aagtcttgat ctcttactc tagattttgc tgcgcg 46

<210> 8
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> attB1

<400> 8
tgaagcctgc ttttttatac taacttgagc gaa 33

<210> 9
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> attB2

<400> 9
ttcgtcaag ttagtataaa aaagcaggct tca 33

<210> 10
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer: attPdwn2

<400> 10
tcttctcggg cataagtcgg acacc 25

<210> 11
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:CMVen

<400> 11
ctcacgggga tttccaagtc tccac 25

<210> 12
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:attPdwn

<400> 12
cagaggcagg gagtgggaca aaattg 26

<210> 13
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:CMVEN2

<400> 13
caactccgcc ccattgacgc aaatg 25

<210> 14
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:L1

<400> 14
agtatcgccg aacgattagc tcttca 26

<210> 15
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:F1 rev

<400> 15
gccgatttcg gcctattggt taaa 24

<210> 16
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:RED

<400> 16
ccgccgacat ccccgactac aagaa

25

<210> 17
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:L2rev

<400> 17
ttccttcgaa ggggatccgc ctacc

25

<210> 18
<211> 22118
<212> DNA
<213> Mus musculus

<300>
<308> GenBank X82564
<309> 1996-04-09

<400> 18
gaattccctt atccctaate cagatttggtg gaataaacttg gtatagatgt ttgtgcatta 60
aaaaccctgt aggatcttca ctctagggtca ctgttcagca ctggaacctg aattgtggcc 120
ctgagtataa ggtcctggga catatgcagt tctgcacaga cagacagaca gacagacaga 180
cagacagaca gacagacgtt acaaacaacac acgttgagcc gtgtgccaac acacacacaa 240
acaccactct ggccataaatt attgaggacg ttgatttatt attctgtgtt tgtgagtcgt 300
tctgtctgtc tgtctgtctg tctgtctgtc tatcaaacca aaagaaacca aacaattatg 360
cctgcctgcc tgcctgcctg cctacacaga gaaatgattt cttcaatcaa tctaaaacga 420
cctcctaagt ttgccttttt tctctttctt tatctttttc ttttttcttt tcttcttctt 480
tcttctcttc cttccttctt tcttctcttt ctttctttct ttctttcttt cttactttct 540
ttctttcttt cttacattta ttcttttcat acatagtttc ttagtgtaag catccctgac 600
tgtcttgaag acactttgta ggccctcaatc ctgtgaagag cttcctctgc ttttcaaatg 660
ctggcatgaa tgttgtaact cactatgacc agcttagtct tcaagtctga gttactggaa 720
aggagtcca aagaagactgg ttatatTTTT catttattat tgcattttta ttaaaattta 780
atttcacca aagaatttag actgaccaat tcagagctcg ccgtttaaaa gcataaggaa 840
aaagtaggag aaaaacgtga ggctgtctgt ggatggctga ggctgcttta gggagcctcg 900
tcaccattct gcacttgcaa accggggccac tagaacccgg tgaagggaga aaccaaagcg 960
accttgaaac aatagggtcac atgaaggcca gccacctcca tcttggtgtg cgaggattca 1020
gttagcagac aagatggctg ccattgcacat gttgtctttc agcttggtga ggtcaaagta 1080
caaccgagtc acagaacaag gaagtataca cagtgagttc caggtcagcc agagtttaca 1140
cagagaaacc acatcttgaa aaaaaacaaa aaataaatta aataaatata atttaaaat 1200
ttaaaaatag ccgggagtgaa tggcgcatgt ctttaatccc agctctcttc aggcagagat 1260
gggaggattt ctgagtttga ggccagcctg gtctgcaaag tgagttccag gacagtccag 1320
gctatacaga gaaaccctgt cttgaaaact aaactaaact aaactaaaaa 1380
aatataaaat aaaaatttta aagaatttta aaaaactaca gaaatcaaac ataagcccac 1440
gagatggcaa gtaactgcaa tcatagcaga aatattatac acacacacac acacagactc 1500
tgtcataaaa tccaatgtgc cttcatgatg atcaaatttc gatagtcagt aatactagaa 1560
gaatcatatg tctgaaaata aaagccagaa ccttttctgc ttttggtttt ttttgcccca 1620
agatagggtt tctctcagtg tatccctggc atccctgect ttttggtttt ttgtaggttt 1680
ggtagcctca aactcagaga ggtcctctct cctctctctg cctgcctgct gcctgcctgc 1740
ctgcctgctt gcctgcctca cttcttctgc caccacacac accgagtcga acctaggatc 1800
tttattttct tctctttctc tcttctttct ttctttcttt ctttctttct tcttctttct 1860
ctttctttct tcttatttca attagttttc aatgtaagtg tgtgtttgtg ctctatctgc 1920
tgcctatagg cctgcttgcc aggagagggc aacagaacct aggagaaacc accatgcagc 1980
tcctgagaat aagtgaaaaa acaacaaaaa aaggaaattc taatcacata gaatgtagat 2040
atatgccgag gctgtcagag tgctttttta ggcttagtgt gctcggcctg aagtaatgaa aattgttgtg 2100
tgtcttttat ccaaacacag acattaatct gctcggcctg catgtctgtt gtctgcagtg 2160
agaccaggct ggccttgaac cccggactga tttcttctt ctttcttctt gcttccctaa tgctgcgatt 2220
aaaggcatgt gccaccactg ctcttctctc tttcttcttt tttttttttt tggaaaatac 2280
ctttctttct ctttttcttt ttttttttaa aatttgctta aggttaaagg tgtgctccac 2340
aattgcctca gctctgctct aattctcttt ttagaagaaa aaacaaaaaa aaaaacaaa 2400
cagtatgtat gtatgtatat ttagaagaaa tactaatcca ttaataactc ttttttctta 2460
aaattcatgt cattcttgtt ccacaaagtg agttccagga cttaccagag aaaccctgtg 2520
2580

ttcaaat	tgtgttcaag	gtcaccctgg	cttacaaagt	gagttccaag	tccgataggg	2640
ctacacagaa	aaaccatatc	tcagaaaaaa	aaaaagttcc	aaacacacac	acacacacac	2700
acacacacac	acacacacac	acacacacac	acacacacag	cgcgcgcgcg	cgatgagggg	2760
aagtcgtgcc	taaaataaat	atTTTTtctgg	ccaaagtga	agcaaatac	tatgaagagg	2820
tactcctaga	aaaaataaat	acaaacgggc	TTTTtaata	ttccagcact	gttttaattt	2880
aactctgaat	ttagtcttgg	aaaagggggc	gggtgtgggt	gagtgagggc	gagcgagcag	2940
acggggcggg	gggcgggtga	gtggccggcg	gcgggtggcg	cgagcaccag	aaaacaacaa	3000
accccaagcg	gtagagtgtt	ttaaaaatga	gacctaaatg	tggtggaacg	gaggtcgccg	3060
ccaccctcct	cttccactgc	ttagatgtct	ccttccccct	actgtgctcc	cttccccctaa	3120
ctgtgcctaa	ctgtgcctgt	tccttcaccc	cgctgattcg	ccagcgacgt	actttgactt	3180
caagaacgat	tttgccctgt	ttcaccgctc	ctgtcatac	tttcgttttt	gggtgcccga	3240
gtctagcccg	ttcgctatgt	tcgggcggga	cgatggggac	cgttttgtgc	actcgggaga	3300
agtggtgggt	gggtacgctg	ctccgtcgtg	cggtgcgtgag	tgccggaacc	tgagctcggg	3360
agaccctccg	gagagacaga	atgagtgaat	gaatgtggcg	gcgcgtgacg	gatctgtatt	3420
ggtttgtatg	gttgatcgag	accattgtcg	ggcgacacct	agtggtgaca	agtttcggga	3480
acgctccagg	cctctcaggt	tggtgacaca	ggagagggaa	gtgctctgtg	tgagcgagcc	3540
agggtgacag	gaggccgggc	aagcaggcgg	gagctgtctg	gagatgggtg	cggttttaag	3600
gacggctctc	aaacaaggagg	tcgtacaggg	agatggccaa	agcagaccga	gttgctgtac	3660
goccttttgg	gaaaaatgct	aggggttggtg	gcaacgttac	taggtcgacc	agaaggctta	3720
agtcctaccc	ccccccccct	TTTTTTTTtt	tttctctcag	aagccctctc	ttgtccccct	3780
caccgggggc	accgtacatc	tgaggccgag	aggacgcgat	gggccccggc	tccaagcccg	3840
tgtggctcgg	ccagctggcg	cttcgggtct	TTTTTTTTtt	TTTTTTTTtt	TTTTcctcca	3900
gaagccttgt	ctgtcgctgt	caccgggggc	gctgtacttc	tgaggccgag	aggacgcgat	3960
gggccccggc	ttccaagcgc	gtgtggctcg	gccagctgga	gcttcgggtc	TTTTTTTTtt	4020
TTTTTTTTtt	TTTTTTTTct	cagaagcctt	gtctgtcgct	gtcaccgggg	gcgctgtact	4080
tctgaggccg	agaggacgcg	atgggtcggc	ttccaagccg	atgtggcggg	gccagctgga	4140
gcttcgggtt	TTTTTTTTtt	ctccagaagc	cctctcttgt	ccccgtcacc	gggggcgctg	4200
tacttctgag	gccgagagga	cgtgatgggc	ccgggttcca	ggcggatgtc	gccccgtcag	4260
ctggagcttt	ggatcttttt	TTTTTTTTtt	cctccagaag	ccctctcttg	tccccgtcac	4320
cgggggcacc	ttacatctga	gggcgagagg	acgtgatggg	tcgggtcttc	aagccgatgt	4380
ggcggggcca	gctggagctt	cgggtttttt	TTTTTctctc	cagaagccct	ctcttgtccc	4440
cgtcaccggg	ggcgtgttac	ttctgaggcc	gagaggacgt	gatgggcccc	ggttccaggt	4500
ggatgtcgcc	cggtcagctg	gagctttgga	tcattttttt	TTTTTctctc	agaagccctc	4560
tcttgtcccc	gtcaccgggg	gcaccgtaca	tctgaggccg	agaggacacg	atgggcctgt	4620
cttccaagcc	gatgtggccc	ggccagctgg	agcttcgggt	cTTTTTTTTt	TTTTTctctc	4680
cagaagccct	gtctgtcgct	gtcaccgggg	gcgctgtact	tctgaggccg	agaggacgcg	4740
atggggcccg	cttccaagcc	ggtgtggctc	ggccagctgg	agcttcgggt	ctTTTTTTTT	4800
TTTTTTTTtt	ttcctccaga	aaacctgtct	gtcgtgtcca	cccggggcgc	ttgtacttct	4860
gatgccgaga	ggacgcgatg	ggcccgctct	ccaggccgat	gtggcccggt	cagctggagc	4920
tttggatctt	TTTTTTTTtt	TTTTTctcca	gaagccctct	cttgtccccg	tcaccggggg	4980
caccttacat	ctgaggccta	gaggacacga	tgggcccggg	ttccaggccg	atgtggcccc	5040
gtcagctgga	gcttttgata	TTTTTTTTtt	TTTTTctcca	gaagccctct	tgtccccgtc	5100
accggtggca	ctgtacatct	gaggcggaga	ggacattatg	ggccccgctt	ccaatccgat	5160
gtggcccggt	cagctggagc	tttggatctt	attTTTTTTTT	taattTTTTt	ttccagaagc	5220
cctcttgtcc	ctgtcacccg	tggcacggta	catctgaggc	cgagaggaca	ttatgggccc	5280
ggcttccagg	ccgatgtggc	ccggtcagct	ggagctttgg	atctTTTTTT	TTTTTTTTtt	5340
TTTTTctctc	agaagccctc	tctgtccctg	tcaccggggg	ccctgtacgt	ctgaggccga	5400
gggaaagcta	tgggcgcggg	tttctttcat	tgacctgtcg	gtcttatcag	ttctccgggt	5460
tgtcagggtc	gaccagttgt	tcctttgagg	tcgggttctt	ttcgttatgg	ggctattttt	5520
gggccacctc	cccaggtatg	acttccaggc	gtcgttgtct	gcctgtcact	ttcctccctg	5580
tctcttttat	gcttgtgata	TTTTctatct	gttccatttg	gacctggaga	taggtactga	5640
cacgctgtcc	TTTTcctatt	aacactaaag	gacactataa	agagaccctt	tcgatttaag	5700
gctgttttgc	ttgtccagcc	tattcttttt	actggcttgg	gtctgtcgcg	gtgcttgaag	5760
ctgtccccga	gccacgcttc	ctgctttccc	gggcttgctg	cttgctgttg	cttgctgtgg	5820
gcagcttgtg	acaactgggc	gctgtgactt	tgctgcgtgt	cagacgtttt	tccccatttc	5880
cccaggttgt	cgtttgtaca	cctgtcccgg	ttggaatggt	ggagccagct	gtgggtgagg	5940
gccaccttat	ttcggctcac	TTTTTTTTtt	TTTTTctctc	ttggagtccc	gaacctccgc	6000
tctttttctc	tccgggtctt	tcttccacat	gcctcccgag	tgcatTTTct	TTTTTTTTtt	6060
ttcttttttt	TTTTTTTTtt	ttggggaggt	ggagagtccc	gagtaactca	ctcctgtctg	6120
tggtgtccaa	gtggttcatg	cacgtgcctc	ccgagtgcac	TTTTTTTTtt	ggcagctcgt	6180
cgttgtgttc	tcttgtttct	tgtctgcccc	tatcagtaac	tgtcttgccc	cgcgtgtaag	6240
acattctcat	ctcgttgtgt	tctcccagatt	gcgcgtcgtt	gctcactctt	agatcgatgt	6300
ggtgtctcgg	agtctctctc	gggcccaggc	caagccgcgc	caggcgaggg	acggacattc	6360
atggcgaaat	gcggccgcct	ttctcgttct	gccagcgggc	cctcgtctct	ccacccccat	6420
cgtctgcggg	tggtgtgtgg	aaggcagggg	tcgggctctc	cggccccgacg	ctgccccgcg	6480
cgcacttttc	tcagttggtc	gcgtgggtcc	tggtggtgtg	tgaggcgccc	ggtttgcgcc	6540
tcacgtgttt	cactttgggtc	gtgtctcgct	tgaccatggt	cccagagtcg	gtggatgtgg	6600

ccggtggcgt	tgcataccct	tcccgctctgg	tgtgtgacag	cgctgtttct	tgtaagcgtc	6660
gaggtgctcc	tgagagcgttc	caggtttgtc	tcctaggtgc	ctgcttctga	gctgggtggg	6720
gcgctcccca	ttccctgggtg	tgccctccggt	gctccgctctg	gctgtgtgcc	ttcccgtttg	6780
tgtctgagaa	gcccgtgaga	gggggggtcga	ggagagaagg	aggggcaaga	cccccttct	6840
tcgtcgggtg	aggcgccac	cccgcgacta	gtacgcctgt	gcgtagggtc	ggtgctgagc	6900
ggtcgcggtc	ggggttggaa	agtttctcga	gagactcatt	gctttcccg	ggggagcttt	6960
gagaggcctg	gctttcgggg	gggaccggtt	gcagggtctc	ccctgtccgc	ggatgctcag	7020
aatgcccttg	gaagagaacc	ttcctgttgc	cgcagacccc	cccgcgcggt	cgcccgcggtg	7080
ttggctctct	ggtttccctg	tgtgctcgtc	gcctgcctcc	tctctcggtg	gcccgggctc	7140
gtcgggggtt	tggttccgct	ccgcccctcag	tgagaaagtt	tccttctcta	gctatcttcc	7200
ggaaagggtg	cgggttctt	acggtctcga	ggggtctctc	ccgaatgggtc	ccctggagggtg	7260
ctcgccccct	gaccgcctcc	cgcgcgcgca	gcgtttgtctc	tctcgtctac	cgccggcccg	7320
ggcctccccg	ctccgagttc	ggggagggtg	cacgcggggg	agagcctgtc	tgctgctcctg	7380
ccgttgctgc	ggagcatgtg	gctcggcttg	tgtggttgg	ggctggggag	agggctccgt	7440
gcacaccccc	gcgtgcgcgt	actttcctcc	cctcctgagg	gcccgcgtgc	ggacgggggtg	7500
tggttaggcg	acggtgggct	cccgggtccc	caccgctctt	cccgtgcctc	accgctgcct	7560
tcgctcgctg	gcgtccctct	cgcctcgcctc	cacgactttg	gcccgtcccg	cgacgggggc	7620
ctgcgcgcgc	cgtggtgctg	gctgtgtgct	tctcgggctg	tgtggttgtg	tcgcctcgcc	7680
cccccttcc	cgcggcagcg	ttcccacggc	tggcgaaatc	gcgggagtc	tccttcccc	7740
cctcgggttc	gagagggtcc	gtgtctggcg	ttgattgatc	tcgctctcgg	ggacggggacc	7800
gttctgtggg	agaacggctg	ttggccgcgt	ccggcgcgac	gtcggacgtg	gggacccact	7860
gcccgtcggg	ggtcttctgc	ggttaggcctc	gggtgtgtcgg	catcggtctc	tctctcgtgt	7920
cgggtctcgc	tcctcgggct	cccggggggc	cgctcgtgtt	cgggtcgggt	cgccgctgca	7980
ggtgtggtgg	gactgctcag	gggagtgggtg	cagtgtgatt	cccgcgggtt	ttgcctcgcg	8040
tgccctgacc	ggtccgacgc	ccgagcggtc	tctcgggtccc	ttgtgaggac	ccccctccgg	8100
gagggggccg	tttcggccgc	ccttgccgct	gtcgcggg	ctcgttctgc	tgtgtcgttc	8160
ccccctcccc	gctcgcgcga	gcccgtcttt	tttctctct	ccccctctct	cctctgactg	8220
accgctggcc	gtgctgtcgg	accccccgca	tggggggcgg	cgggcacgta	cgcgtccggg	8280
cggctacogg	ggtcttgggg	ggggggccgag	gggtaagaaa	gtcggctcgg	cgggcgggag	8340
gagctgtggg	ttggaggggc	tcccggcccc	gcccgcgtgg	cggtgtcttg	cgcggctctg	8400
gagagggtcg	cgtgcgaggg	gaaaagggtg	ccccgcgagg	gcaaagggaa	agaggctagc	8460
agtggctcatt	gtcccagcgg	tgtggtgggt	gtgtggccga	gggtgcgtctg	gggggctcgt	8520
ccggccctgt	cgtccgctcg	gaaggcgcg	ggtggggcct	gcccggagtgc	cgagggtgggt	8580
accctggcgg	tgggattaac	cccgcgcgcg	tgtcccgggt	tggcggtggg	gggtccgggtc	8640
gatgtctacc	tcctctcccc	cgaggctctca	ggccttctcc	gcgcgggctc	tcggccctcc	8700
cctcgcttcc	ccctctcgcg	gggttcaagt	cgctcgtcga	cctccccctc	tcggtccttc	8760
catctctcgc	gcaatggcgc	cgcccagatt	cacggtgggt	tcgtcctccg	cctccgcttc	8820
tcgcgggggg	ctggccgctg	tcgggtctct	cctgcgcgac	ccccgttggc	gtggtcttct	8880
ctcgccggct	tcgcgggactc	ctggcttcgc	cggagggttc	agggggcttc	ccgggtcccc	8940
gacgttgccg	ctcgctgctg	tgtgcttggg	ggggggccgc	tgccggcctc	gcccgcgcgt	9000
gagccccctg	cgcacccgc	ggtgtgcggt	ttcgcgcgcg	ggtcagttgg	gcccgtggct	9060
tggtgcgctg	cgggagcgtg	tcgcgcctgc	ggcggttaga	cgcggtgtgc	gcccgggtcc	9120
gacgggtggc	ctatccaggg	ctcgcccccg	ccgaaccccc	cctgcccgtc	ccggtgggtg	9180
tcgttgggtg	ggggagtga	tggtgtctacc	ggctattccc	tcgcgcgtgg	tttgactgtc	9240
tcgcgggtgt	cgcgcttctc	tttcgcgcaa	ccccacgcgc	aaccacccac	cctgctctcc	9300
cggcccggtg	cggtcgacgt	tcgggtctct	ccgatgccga	ggggttcggg	atttgtgcgc	9360
gggacggagg	ggagagcggg	taagagaggt	gtcggagagc	tgtcccgggg	cgacgctcgg	9420
ggtggctttg	ccgcgtgcgt	gtgctcgcgg	acgggttttg	tcggaccccg	acggggtcgg	9480
tcggcccgca	tgcactctcc	cgttcgcgcg	gagcgcgcgc	ccggctcacc	cccgtttgt	9540
cctcccgcga	ggctctccgc	cgcgcgcgcg	tcctcctcct	ctctcgcgct	ctctgtcccg	9600
cctggtcctg	tcacaccccc	gacgctccgc	tcgcgcttcc	ttacctgggt	gatcctgcca	9660
ggtagcatat	gcttgtctca	aagattaaag	catgcatgtc	taagtacgca	cggccgggtac	9720
agtgaactg	cgaatggctc	attaaatcag	ttatggttcc	tttgggtcgt	cgctcctctc	9780
ctacttggat	aactgtggta	attctagagc	taatacatgc	cgacggggcg	tgacccccct	9840
tcgggggggg	ggatgctgtc	atttatcaga	tcaaaaccaa	cccggtgagc	tccttcccgg	9900
ctccggccgg	gggtcggg	ccggcggtt	ggtgactcta	gataacctcg	ggccgatcgc	9960
acgccccccg	tgccggcgac	gacccattcg	aacgtctgcc	ctatcaactt	tcgatggtag	10020
tcgcgctgcc	taccatgggtg	accacgggtg	acggggaatc	aggggttcgat	tcgggagagg	10080
gagcctgaga	acaggctacc	acatccaagg	aaggcagcag	gcgcgcaaat	taccactcc	10140
cgacccgggg	aggtagtac	gaaaaataac	aatacaggac	tctttcgagg	cctgtatatt	10200
ggaatgagtc	cactttaaat	cctttaacga	ggatccattg	gagggcaagt	ctggtgccag	10260
cagcccggtg	aattccagct	ccaatagcgt	atattaaagt	tgctgcagtt	aaaaagctcg	10320
tagttggatc	ttggggcggt	gcggcggtgc	cgcgcgagg	cgagtacccg	ccggtccccg	10380
cccccttgct	ctcggcgccc	cctcgatgct	ccttagctgag	tgccccgcgg	ggcccggaagc	10440
ggttactttg	aaaaaattag	agtgttcaaa	gcaggccgca	gcccgcctgga	taccgcagct	10500
aggaataatg	ggataggacc	gcggttctat	ttgttgggt	tcgggaactg	aggccatgat	10560
taagaggggac	ggccgggggg	attcgtattg	cgcgcgtaga	ggtgaaattc	ttggaccggc	10620

gcaagacgga	ccagagcgaa	agcatttggc	aagaatgttt	tcattaatca	agaacgaaag	10680
tcggaggttc	gaagacgatc	agataccgtc	gtagttccga	ccataaacga	tgccgactgg	10740
cgatgcggcg	gcgttattcc	catgacccgc	cgggcagctt	ccgggaaacc	aaagtctttg	10800
ggttccgggg	ggagtatggt	tgcaaagctg	aaacttaaa	gaattgacgg	aagggcacca	10860
ccaggagtgg	gcctgcggct	taatttgact	caacacggga	aacctcacc	ggcccggaca	10920
cggacaggat	tgacagattg	atagctcttt	ctcgattccg	tgggtggtgg	tgcatggccg	10980
ttcttagttg	gtggagcgat	ttgtctggtt	aattccgata	acgaacgaga	ctctggcatg	11040
ctaactagtt	acgcgacccc	cgagcggtcg	gcgtccccc	acttcttaga	gggacaagtg	11100
gcgttcagcc	acccgagatt	gagcaataac	aggtctgtga	tgcccttaga	tgtccggggc	11160
tgacgcgcg	ctacactgac	tggtcagcgg	tgtgcctacc	ctgcgcggc	aggcgcgggt	11220
aaaccgttga	acccatttcg	tgatggggat	cggggattgc	aattattccc	catgaacgag	11280
gaattcccag	taagtgcggg	tcataagctt	gcgttgatta	agtccctgcc	ctttgtacac	11340
accgcggctc	gctactaccg	attggatggt	ttagtgaggc	cctcggatcg	gccccgcgg	11400
ggtcggccca	cgccctggc	ggagcgctga	gaagacggtc	gaacttgact	atctagagga	11460
agtaaaagtc	gtaacaaggt	ttccgtaggt	gaacctgcgg	aaggatcatt	aaacgggaga	11520
ctgtggagga	gcggcggcgt	ggcccgctct	ccccgtcttg	tgtgtgtcct	cgccgggagg	11580
cgctgcgtc	ccgggtcccg	tcgcccgcgt	tcgggacgag	gtgtctggag	tgaggtgaga	11640
gaagggtggg	gtggggtcgg	tctgggtccg	tctgggaccg	cctccgattt	ccccccccc	11700
tccctctccc	ctcgctcggc	cttgacctcg	ccacctacc	gcggcggcgg	ctgctcggc	11760
gcgtcttggc	tctttcccg	ccgctcttc	cgctgtctac	aggggcggta	cgctgttacg	11820
ggtttttgac	ccgtcccggt	ggcgttcgg	cgctcggggc	cgcgctttgc	tctcccgga	11880
cccatcccc	ccgcggctct	ggcttttcta	cgctggctgg	ggcgggtgtc	gcgtgtgggg	11940
ggatgtgagt	gtcgcgtgtg	ggctcgcccg	tcccgatgcc	acgcttttct	ggcctcgct	12000
gtccctcccc	ctccgtgtcc	gggtacctag	ctgtcgctgt	ccggcgcgga	ggtttaagga	12060
ccccgggggg	gtcgccctgc	cgcccccagg	gtcggggggc	gggtggggccc	gtagggaagt	12120
cggctcgttc	ggcggctctc	cctcagactc	catgaccctc	ctccccccg	tgccgcgct	12180
cccgaggcgg	cggtcgtgtg	gggggggtga	tgtctggagc	ccccctcggg	gcccgtgggg	12240
cccgacccgc	gcgcgcggct	tgccccgatt	ccgcgggtcg	gtcctgtcgg	tgccggtcgt	12300
gggttcccg	gtcgttcccc	tgtttttccg	ctccccgacc	tttttttttc	ctcccccca	12360
caactgtctc	gtttcgttcc	tgttggccgg	cctgaggtcta	ccccctcggt	catctgttct	12420
cctctctctc	cggggagagg	agggcggtgg	tgttgggggg	actgtgccgt	cgctagcacc	12480
cgtgagttcg	ctcacaccgc	aaataccgat	acgactctta	gcggtggatc	actcggctcg	12540
tgctgcgatg	aagaacgcag	ctagctgcga	gaattaatgt	gaattgcagg	acacattgat	12600
catcgacact	tcgaacgcac	ttgcggcccc	gggttccctc	cgggggtacg	cctgtctgag	12660
cgtcggttga	cgatcaatcg	cgtcacccgc	tgcggtgggt	gctgcgcggc	tgggagtttg	12720
ctcgcagggc	caacccccca	acccgggtcg	ggccctccgt	ctcccgaaat	tcagacgtgt	12780
gggcgggtgt	cggtgtggcg	cgccgcggcg	cgctcgcggg	cctgggtctc	cccgcccatc	12840
cgcgctcgcg	gcttcttccc	gctccgccgt	tccccccctc	gcccgtgcac	cccggctcct	12900
gcctcgcgtc	gggcgcctcc	ctcaccaagt	ctcaccaagt	tttctcgggt	ccgtgccccg	12960
tgggaaccca	ccgcgccccc	gtggcgcccg	ggggtggggc	cgctccgcat	tgctctgggt	13020
gaggttggcg	gttgagggtg	tgcgtgcgcc	gaggtgggtg	tccgtccctc	gcccgcggcg	13080
ggctgtcggg	gtggcgctcg	acgagggcgg	ctcggtcgcc	tgccgtgtgt	gtctgtgtgt	13140
gtttgggtct	tgcgctgggg	gagggcgggt	cgaccgctcg	cgggggtggc	gcggtcgccc	13200
ggcgcccgcg	acccctcggc	ttgtgtggag	ggagagcgag	ggcgagaacg	gagagaggtg	13260
gtatccccgc	tgccgttgcg	agggaggggt	tggcgtccc	cgctccgtcc	tccctccctc	13320
cctcgggtgg	cgcccttcgg	ccgcacgcgg	ccgctagggg	cggtcggggc	ccgtggcccc	13380
cgtggctctt	cttcgtctcc	gcttctctct	cacccggggc	gtacccgctc	cggcgccggc	13440
ccgcgggacg	cccgggcgtc	cgtgcgcgga	tgcgagtcac	ccccgggtgt	tgcgagttcg	13500
gggagggaga	gggcctcgct	gacccggtgc	gtcccggctt	ccctgggggg	gacccggcgt	13560
ctgtgggctg	tgcgtcccg	gggttgcgtg	tgagtaagat	cctccacccc	cgccgcccct	13620
ccctcccgcc	ggcctctcgg	ggacccccct	agacgggtcg	ccggctcgtc	ctcccggtgc	13680
gcggggtgoc	gtctctttcc	cgcccgccct	ctcgtctctc	tcttcccccg	gctggggcgc	13740
tgtccccctc	ttctgaccgc	gacctcagat	cagacgtggc	gacccgctga	atttaagcat	13800
attagtgcgc	ggaggaaaag	aaactaacca	ggattccctc	agtaacggcg	agtgaacagg	13860
gaagagccca	gcgcggaatc	cccgcccgcg	gtcgcggcgt	gggaaatgtg	gcgtacggaa	13920
gacccactcc	ccggcgccgc	tgcgtggggg	cccaagtcct	tctgatcgag	gcccagcccc	13980
tggacgggtg	gagggcggta	gcggccccgg	cgcgccgggc	tccgggtctc	ccggagtcgg	14040
gttgcttggg	aatgcagccc	aaagcgggtg	gtaaactcca	tctaaggcta	aataccggca	14100
cgagaccgat	agtcaacaag	taccgtaagg	gaaagttaga	aagaactttg	aagagagagt	14160
tcaagagggc	gtgaacccgt	taagaggtaa	acgggtgggg	tccgcgcagt	ccgcccggag	14220
gattcaaccc	ggcgggcgcg	gtccggccgt	gcccgggtgt	cccggcggat	ctttcccgct	14280
ccccgttctc	cccgacccct	ccaccccgcg	gtcgttcccc	tcttctctcc	cgcgtccggc	14340
gcctccggcg	ggggcgcgcg	gggggtggtg	gggtggtggc	cgccggggcg	gcccgggggt	14400
gggtcggcgg	ggggccggcg	ccggccggcg	accggccggc	gcccggggcg	cttccaccgt	14460
ggcggtgcgc	cgcgaccggc	tccgggacgg	ccgggaaggg	ccgggtgggga	aggtggctcg	14520
ggggggggcg	cgcgctctcag	ggcgcgccga	accacctcac	cccgagtgtt	acagccctcc	14580
ggcgcgcgtt	tcgcggaatc	ccggggccga	ggaagccaga	taccgctcgc	cgcgctctcc	14640

ctctcccccc	gtccgcctcc	cgggcgggcg	tgggggtggg	ggccggggccg	ccccccccc	14700
ggcgcgaccc	ctctcccccc	cccccccgct	gcctctctcg	gggcccgggtg	ggggggcgggg	14760
cggactgtcc	ccagtgccgc	ccgggcgctc	tccgcgcctc	gggtcccggg	gggaccgtcg	14820
gtcacgcgtc	tcccgcagaa	gccgagcgca	cggggtcgge	ggcgatgtcg	gctacccacc	14880
cgaccgcgtc	tgaaacacgg	accaaggagt	ctaaccgcgtg	cgcgagtcag	gggctcgctc	14940
gaaagccgcc	gtggcgcaat	gaaggtgaag	ggcccccgccc	ggggggcccg	ggtgggatcc	15000
cgaggcctct	ccagtcgcgc	gagggcgcc	caccggcccc	tctcgcccc	cgcgccgggg	15060
agggtggagca	cgagcgtaag	cggttaggacc	cgaaagatgg	tgaactatgc	ttgggcaggg	15120
cgaagccaga	ggaaactctg	gtggaggtcc	gtagcggtcc	tgacgtgcaa	atcggtcgct	15180
cgacctgggt	ataggggcca	aagactaatc	gaacctcta	gtagctgggt	ccctccgaag	15240
tttccctcag	gatagctggc	gctctcgctc	cgacgtacg	cagttttatc	cggtaaagcg	15300
aatgattaga	ggtcttgggg	ccgaaacgat	ctcaacctat	tctcaaaactt	taaagtgggt	15360
agaagcccg	ctcgctggcg	tggagccggg	cgtggaatgc	gagtgcctag	tgggcccactt	15420
ttggttaagca	gaactggcgc	tgcgggatga	accgaacgcc	gggttaaggc	gcccgatgcc	15480
gacgtcatc	agaccccgca	aaaggtgttg	ggtgatatag	acagcaggac	ggtggccatg	15540
gaagtccgaa	tccgctaagg	agtgtgtaac	aactcacctg	ccgaatcaac	tagccctgaa	15600
aatggatggc	gctggagcgt	cgggcccgta	cccgcccgct	gccgcagtcg	gaacggaaacg	15660
ggacgggagc	ggccgcgggt	gcgcgtctct	cggggtcggg	ggtgcgtggc	gggggcccgt	15720
cccccgctc	ccctccgcgc	gcccgggtcg	cccccgccgc	gtcgggcccc	gcccagccta	15780
cgccgcgacg	agtgcgcgtt	ccgctgcgtg	gagccttgaa	gcctaggggcg	cgggcccggg	15840
tggagccgoc	gcaggtgcag	atcttggtgg	tagtagcaaa	tattcaaacg	agaactttga	15900
aggccgaagt	ggagaagggt	tccatgtgaa	cagcagttga	acatgggtca	gtcggctcctg	15960
agagatgggc	gagtgcgcgtt	ccgaagggac	ggcgatggc	ctccgttgcc	ctcgcccgat	16020
cgaaaggggag	tccgggttcag	atccccgaat	ccggagtggc	ggagatgggc	gccgcgaggc	16080
cagtgcggta	acgcgaccca	tcccggagaa	gcccggcgga	ggcctcgggg	agagtctctc	16140
tttctttgtg	aaggggcagg	cgccctggaa	tgggttcgcc	ccgagagagg	ggcccggtgc	16200
ttggaaagcg	tccgggttcc	ggcgccgctc	ggtgagctct	cgctggccct	tgaaaatccg	16260
ggggagaggg	tgtaaatctc	gcgcggggcc	gtacccatat	ccgcagcagg	tctccaaggt	16320
gaacagcctc	tggcatgttg	gaacaatgta	ggtaagggaa	gtcggcaagc	cggtaccgt	16380
acttcgggat	aaggattggc	tctaagggtc	gggtcggtcg	ggctggggcg	cgaagcgggg	16440
ctggggcgcg	gcccgggctg	gacgagggcg	cgccgcctcc	tcccacgtcc	gggggagacc	16500
cccgtccttt	ccgcctcggc	ccgcctccc	ctctccccg	cggggccccg	tccgtccccg	16560
cgctcgctgc	acctctcttc	ccccctcctt	cttcccgctg	gggggcccgg	cggggggtcg	16620
cgcgcggcgc	gggctccggg	gcccgggggt	caaccccccg	gggggtcccg	agccgggagga	16680
accagcggtc	cccggtgggg	cgggggggcc	ggacactcgg	ggggccggcg	gcggccggca	16740
ctctggacgc	gagccggggc	cttcccggtg	atcgctcag	ctgcggcggg	cgctcgcgcc	16800
gctcccgggg	agcccgggcg	gtgcccggcg	gggtcccctc	cccgcggggc	ctcgctccac	16860
ccccccatcg	cctctcccga	ggtgcgtggc	gggggcccgc	gggcgtgtcc	cgcgctgtg	16920
gggggaacct	ccgcgtcggt	gttccccgc	gggttcggcc	ccccggggcg	cggttttccg	16980
cgcgccgccc	ccgcctcggc	cgggcgctag	cagccgactt	agaactgggtg	cggaaccagg	17040
gaatccgact	gtttaattaa	aacaaagcat	cgcgaaggcc	cgccgcccgg	ggtgacgca	17100
tgtgatttct	gcccagtgat	ctgaatgtca	aggaagaaga	attcaatgaa	gcgcgggttaa	17160
acggcgggag	taactatgac	tctcttaagg	tagccaaatg	cctcgctcatc	taattagtga	17220
gcgcgatgaa	tggatgaacg	agattcccac	tgtccctacc	tactatccag	cgaaacacac	17280
gccaagggaa	cggtctggc	ggaatcagcg	agggcggggg	acctgttgga	gcttgactct	17340
agtctggcac	ggtgaagaga	catgagaggt	gtagaataag	tgggaggccc	ccggcgccc	17400
gccccgctct	cgcgctcggg	tccggggcac	ccggcctcgc	gggcccggcg	tgaaatacca	17460
ctactctcat	cggtttttca	ctgaccgggt	gagggcgggg	ggcgagcccc	gaggggctct	17520
cgcttctggc	gccaagcgct	cgctcccgcg	gtgcccggcg	gcgcgacccc	ctccggggac	17580
agtgccaggt	ggggagtttg	actggggcgg	tacacctgtc	aaacggtaac	gcaggtgtcc	17640
taaggcgagc	tcagggagga	cagaaacctc	ccgtggagca	gaagggcaaaa	agctcgcttg	17700
atcttgattt	tcagtaacgaa	tacagaccgt	gaaagcgggg	cctcacgatc	cttctgacct	17760
tttgggtttt	aagcaggagg	tgtcagaaaa	ggtaccacag	ggataactgg	cttgtggcg	17820
ccaagcgttc	atagcgacgt	cgctttttga	tccctcgatg	tcggctcttc	ctatcattgt	17880
gaagcagaat	tcaccaagcg	ttggatttgt	cacccactaa	tagggaacgt	gagctgggtt	17940
tagaccgtcg	tgagaaggt	tagttttacc	ctactgatga	tgtgttgttg	ccatggtaat	18000
cctgctcagt	acgagaggaa	ccgcaggttc	agacatttgg	tgtatgtgct	tggctgagga	18060
gccaatgggg	cgaagctacc	atctgtggga	ttatgactga	acgcctctaa	gtcagaatcc	18120
gcccgaagcg	aacgatacgg	cagcgccgaa	ggagcctcgg	ttggcccccg	atagccgggt	18180
ccccgctcgt	cccgctcggc	ggggtccccc	cggtcccccg	cggcggcgcg	gggtctcccc	18240
ccgcggggcg	tccggacccg	ggtccgggtc	ggagagccgt	tccgtcttggg	aaacgggggtg	18300
cgggcggaaaa	gggggcccgc	ctctcgcccc	tcacgttgaa	cgcacgttcg	tgtggaacct	18360
ggcgctaaac	cattcgtaga	cgacctgctt	gtttcgtagc	gtttcgtagc	tagcagagca	18420
gctccctcgc	tgcgatctat	tgaagtcag	ccctcgacac	aagggtttgt	ctctgcgggc	18480
tttcccgctc	cagccccgct	cgctcgacgc	cgacgtgtc	gcccggcggg	cgctcagggg	18540
gggtcgccct	cgcccccg	cggttgccc	gaacgacctg	gtggtggttg	ggggggggat	18600
cgtcttctcc	tccgtctccc	gaggacgggt	cgtttctctt	tccccctccg	tccgtctcct	18660

tgggtgtggg	agcctcgtgc	cgtcgcgacc	gcgccctgcc	gtcgccctgcc	gccgcagccc	18720
cttgccctcc	ggccttggcc	aagccggagg	gaggaggagg	gggatcgccg	gccggcgccga	18780
ccgcggcgcg	gtgacgcacg	gtgggatccc	catcctcgcc	gggtccgctc	gggacggccg	18840
gttggagggg	cgggaggggg	ttttcccggt	aacgcccgcg	tcggcgccag	gcctctggcg	18900
gccggggggg	cgctctctcc	gcccgcgcat	ccccactccc	gccccctccc	ttcgcgccgc	18960
gccccggcga	cggtgcgtacg	aggggaggat	gtcgcggtgt	ggaggcgagg	aggggtccggc	19020
gccccgcctc	ttccattttt	ttccccccaa	cttcggagggt	cgaccagtag	ttccggcgac	19080
actttgtttt	ttttttttcc	cccgatgctg	gaggtcgacc	agatgtccga	aagtgtcccc	19140
cccccccccc	ccccccggcg	cgagcgccgc	gggcccactct	ggactctttt	tttttttttt	19200
tttttttttt	ttaaattcct	gggaacctta	gggtcgaccag	ttgtccgtct	tttactcctt	19260
catataggtc	gaccagtagt	ccgggtggta	ctttgtcttt	ttctgaaaat	cccagaggtc	19320
gaccagatat	ccgaaagtcc	tctctttccc	tttactcttc	cccacagcga	ttctcttttt	19380
tttttttttt	tttgggtgtg	ctcttttttg	cttatataca	tgtaaatagt	gtgtacgttt	19440
atatacttat	aggaggagggt	cgaccagtag	tcggggcgac	actttgtttt	tttttttttt	19500
tcacccgatg	atggagggtcg	accagatgtc	cgaaagtgtc	ccgtcccccc	cctccccccc	19560
ccgcgcgcgc	gccccgtcac	tctgggactct	tttttttttt	tttttttttt	tttaaatttc	19620
tggaacctta	aggtcgacca	gttgtccgtc	tttccactcat	tcataatagg	cgaccgggtg	19680
tactttgtct	ttttctgaaa	atcgccagag	tcgaccagat	gtcagaaagt	ctgggtggctg	19740
ataaattatc	tgatctagat	ttgtttttct	gtttttccagt	tttgtgttgt	tttgtgttgt	19800
tttgtgttgt	tttgttttgt	tttgttttgt	tttgttttgt	tttgttttgt	tttgttttgt	19860
tttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	19920
gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	19980
tttgtgttgt	tttgtgttgt	tttgtgttgt	tttgtgttgt	tttgtgttgt	tttgtgttgt	20040
tacacaaaca	tgcaactttt	ttaaaaataa	tttttaaaat	aaatgcgaaa	atcgaccaat	20100
tatccctttc	cttctctctc	ttttttaaaa	attttctttg	tggtgtgtgt	tggtgtgtgt	20160
tggtgtgtgt	tggtgtgtgt	tggtgtgtgt	tggtgtgtgt	tggtgtgtgt	tggtgtgtgt	20220
taactataat	aatagggtcg	ccgggtgggt	ccgggtgggt	ccgggtgggt	ccgggtgggt	20280
gcagacttct	gagttcgagg	ccagcctggg	ctacagagga	accctgtctc	gaaaaatgaa	20340
aataaataca	tacatacata	catatacata	catatacata	catatacata	catatacata	20400
gttgaccagt	tgtaaatcct	ttagaatttt	gttttttaatt	aatgtgatag	agagatagat	20460
aatagataga	tggtatagat	gatacaataa	taggtttttt	tttcagtaaa	tatgagggtg	20520
attaaccact	tttccctttt	taggtttttt	ttttttttcc	ctgtccatgt	gggtgtgggt	20580
atttgaaact	aggacccctg	cagggtcaact	ggaaaaacgt	ttttctatat	atataaatag	20640
tggtctgtct	gctgttttgt	tggtttgctt	ctgtgtgtgt	tggtttgctt	ctgtgtgtgt	20700
tggttttttt	tttctcttga	gacagtattt	ctctgtgtgt	cctgggtgcc	tgaaactcac	20760
tctgttagac	agcctggcct	caatcgaaat	cagaaatcct	cctgcctctt	gtctacctcc	20820
caatttttga	gtaaagggtg	gctacaccac	tgctgggcct	tattatcatt	atcattatta	20880
attttattat	tagacagaac	gaaatcaact	agttgggtct	gtttcgttaa	ttcatttgaa	20940
attagtttga	ccaattagtt	ggctgggttt	ggaggtttct	tttgtttccg	atttgggtgt	21000
ttgtggggct	ggggatcagg	tatctcaacg	gaatgcattg	aggttaagggt	gagatggctc	21060
gattttttga	aagattactt	ttcttagtct	gaggaaaaaa	taaaataata	ttgggctacg	21120
tttcattgct	tcattttctat	ttctctttct	tttttttttt	ctttcagata	aggagggtcg	21180
ccagttctct	ctgccttctg	gaagatgtag	gcattgcatt	gggaaaagca	ttgttttgaga	21240
gatgtgctag	tgaaccagag	agttttggat	tcaagccgta	taatgtttat	tacaatatag	21300
aaaagtctta	acaaagtgat	ctttaacttt	tttttttttt	tttctccttc	tacttctact	21360
tggtctcact	ctgccaccaa	cgcgctttgt	acattgaatg	tgagctttgt	tttgcttaac	21420
agacatatat	tttttctttt	ggtttttgct	gacatgggtt	ccctttctat	ccgtgcaggg	21480
ttcccagacg	gccttttgag	aataaaaatg	gaggccagaa	ccaaagtctt	ttgaataaag	21540
caccacaact	ctaacctgtt	tggtgttttt	ccttcccaag	gcacagatct	ttcccagcat	21600
ggaaaagcat	gtagcagttg	taggacacac	tagacgagag	caccagatct	cattgtgggt	21660
ggttgtgaac	caccacccat	gtggttgcct	gggatttgaa	ctcaggatct	tcagaagacg	21720
agtcagggct	ctaaaccgat	gagccatctc	tcagcccttc	ctacattcct	tcttaaggca	21780
tgaatgatcc	cagcatggga	agacagctct	ccctctttgt	ggtatatcac	catatactca	21840
ataaaaataa	gaaatgaatg	aagtctccac	gtattttatt	cttcgagcta	tctaaattct	21900
ctcacagcac	ctccccctcc	cccacactgc	ctttctccct	atgtttgggt	ggggctgggg	21960
gaggggtggg	gtggggggcag	ggatctgcct	gtcttcttgc	aggtctgtga	actatttgcg	22020
atggcctggg	tctctgaact	gttgagcctt	gtctatccag	aggctgactg	gctagttttc	22080
tacctgaagt	ccctgagtag	tgattttcct	gtgaattc			22118

<210> 19
 <211> 175
 <212> DNA
 <213> Mus musculus

<400> 19	ctccccgcgc	gcccccggtg	tcgcccgttc	cgtggcgccg	acaatgcggt	tggtgcgtcca	60
	cgtgtgctgt	tccgtgcagt	gccgttgtgg	agtgcctcgc	tctcctcttc	ctccccggca	120

gcgttccac ggttggggac caccggtgac ctgcacctct tcgggcctgg atccg

175

<210> 20
<211> 755
<212> DNA
<213> Mus musculus

<400> 20	ggtctggtg	gaattgttga	cctcgctctc	gggtgcggcc	tttggggaac	ggcgggggtcg	60
	gtcgtgccc	gcgcgggac	tgtgtcggg	cccacttccc	gctcgagggt	ggcgggtggcg	120
	gcggcggtg	tagtctccc	tgttgctct	tcccgggctc	ttgggggggg	tgccgtcgtt	180
	ttcggggcc	gcgttgctt	gcttacgcag	gcttggtttg	ggactgcctc	aggagtcgtg	240
	ggcgggtgt	ttcccgcgg	ttttgcctcg	cgctgcctg	ctttgcctcg	ggtttgcttg	300
	gttcgtgtc	cgggagcgg	gggttttttt	tttttcgggt	cccggggaga	gggggtttttc	360
	cgggggacgt	tcccgtcgcc	ccctgccggc	gggtgggtttt	cgtttcgggc	tgtgttcgtt	420
	tccccttccc	cgtttcgcgg	tcggttctcc	ccggtcggtc	ggccctctcc	ccggtcggtc	480
	gcccggcccg	gctgcggac	cccccttct	gggggggatg	cccgggcacg	cacgcgtccg	540
	ggcggccact	gtggtccgg	agctgctcgg	caggcgggtg	agccagttgg	aggggcgtca	600
	tgcccccgcg	ggctcccgtg	gccgacgcgg	cggtgttctt	gggggggcct	gtgcgtgcgg	660
	gaaggctcgc	cacgttgtcg	gtccttgcca	gggaaagagg	cttttttttt	ttaggggggtc	720
	gtccttcgtc	gtcccgtcgg	cggtggatcc	ggcct			755

<210> 21
<211> 463
<212> DNA
<213> Mus musculus

<400> 21	ggccgaggtg	cgtctgcggg	ttggggctcg	tccggccccg	tcgtcctccg	ggaaggcgtt	60
	tagcgggtac	cgtgcgcgg	ccgaggtggg	cgcacgtcgg	tgagataacc	ccgagcgtgt	120
	ttctggttgt	tggcgccggg	ggctccggtc	gatgtcttcc	cctccccctc	tcccgcaggc	180
	caggtcagcc	tccgcctgtg	ggcttcgtcg	gccgtctccc	ccccctcac	gtccctcggc	240
	agcgagcccg	tccgttcgac	cttccttccg	ccttcccccc	atctttccgc	gctccgttgg	300
	ccccggggtt	ttcacggcgc	ccccacgct	cctccgcctc	tccgcccggt	gtttggacgc	360
	ctgggtccgg	tctccccgcc	aaaccccggt	tgggttggtc	tccggccccg	gcttgctctt	420
	cgggtctccc	aacccccggc	cggaagggtt	cgggggttcc	ggg		463

<210> 22
<211> 378
<212> DNA
<213> Mus musculus

<400> 22	ggattcttca	ggattgaaac	ccaaaccggt	tcagtttctt	ttccggctcc	ggccggggggg	60
	ggcgcccccg	ggcggttttg	tgagtttagt	aacctcgggc	cgatcgcacg	ccccccgtgg	120
	cggcgacgac	ccattcgaac	gtctgcctta	tcaactttcg	atggtagtcg	atgtgcctac	180
	catggtgacc	acgggtgaag	gggaatcagg	gttcgattcc	ggagaggagg	cctgagaaac	240
	ggctaccaca	tccaaggaag	gcagcaggcg	cgcaaattac	ccactccga	ccccggggagg	300
	tagtgacgaa	aaataacaat	acaggactct	ttcgaggccc	tgtaatggga	atgagtcac	360
	tttaaatcct	ttaagcag					378

<210> 23
<211> 378
<212> DNA
<213> Mus musculus

<400> 23	gatccatttg	agggcaagtc	tgggtgccagc	agccgcggta	attccagctc	caatagcgta	60
	tattaaagtt	gctgcagtta	aaaagctcgt	agttggatct	tgggagcggg	cgggcggtcc	120
	gccgcgaggc	gagtcacgc	ccgtccccgc	cccttgccct	tcggcgcccc	ctcgatgctc	180
	ttagctgagt	tgtcccgcgg	ggcccgaagc	gtttactttg	aaaaaattag	agttgtttca	240
	aagcaggccc	gagccgcctg	gataccgcc	gctaggaaat	aatggaatag	gaccgcggtt	300
	cctattttgt	ttggttttct	gaactgagcc	catgattaag	ggaaacggcc	gggggcattc	360
	ccttattgct	ccccctta					378

<210> 24
<211> 719

<212> DNA
<213> Mus musculus

<400> 24
ggatctttcc cgctccccgt tcctccccgc cctccacccc ggcggtctcc ccccttcttt 60
tccccctctc ggaggggggg gaggtggggg cgcgtggggc gggtcggggg tggggtcggc 120
gggggacccg ccccgcccg gctacgagac ggctgggaag gcccgcggg gaatgtggct cggggggggc 180
ggcgcgctctc agggcgccgc gaaccacctc accccgagtg ttacagccct ccggcccgcc 240
tttcgcgga tccccggggc gaggggaagc ccgatacccg tcgcccgcgt tttccccctc 300
ccccgtccgc ctccccggcg ggcgtggggg tggggggcgg gccgccccct ccacgcccgt 360
ggttttctctc tctccccggtc tcggccgggt tggggggggg agcccgggtg ggggccccgg 420
ggactgtcct cagtgcgccc cgggctcgt cgcgcgctcg ggccccgggg gtctctctcg 480
tcacgcccgc cccgacgaag ccgagcgcac ggggtcggcg gcgatgtcgg ctaccacccc 540
gacccgtctt gaaacacgga ccaaggagtc taacgcgtgc gcgagtcagg ggctcgcacg 600
aaagccgccc tggcgcaatg aaggtgaagg gccccgtccg gggggcccgag gtgggatcc 660
719

<210> 25
<211> 685
<212> DNA
<213> Mus musculus

<400> 25
cgaggcctct ccagtcgcc gagggcgcac caccggcccc tctcgccgc cgcgtcgggg 60
aggtggagca cgagcgtacg cgtaggacg cgaagatgg tgaactatgc ctgggcaggg 120
cgaagccaga ggaactctg gtggaggtcc gtacgggtc tgacgtgcaa atcggtcgtc 180
cgacctgggt atagggcgga aagactaatc gaaccatcta gtagctgggt cctccgaag 240
tttccctcag gatagctggc gctctcgcaa ccttcggaag cagttttatc cgggtaaagg 300
cggaatggat taggaggtct tggggccgga aacgatctca aactatttct caaactttaa 360
atgggtaagg aagcccggct cgctggcgtg gagccggggc tgggaatgca gtgcctagt 420
ggccactttt ggtaagcaga actggcgctg cgggatgaac cgaacgcccg gtttaaggcg 480
ccgatgccga cgctcatcag accccagaaa aggtgttggg tgatatagac agcaggacgg 540
tggccatgga agtcggaatc cgctaaggag tgtgtaacaa ctacactgcc gaatcaacta 600
gccctgaaaa tggatggcg tggagcgtcg gggccatacc cggccgctcg cggcagtcgg 660
aacgggacgg gacgggagcg gccgc 685

<210> 26
<211> 5162
<212> DNA
<213> Artificial Sequence

<220>
<223> Chimeric bacterial plasmid

<400> 26
gacggatcgg gagatctccc gatccccctat ggtcgactct cagtacaatc tgctctgatg 60
ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg 120
cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc 180
ttaggggttag gcgtttttgcg ctgcttcgcg atgtacggggc cagatatacg cgttgacatt 240
gattattgac tagttattaa tagtaataca ttacgggggtc attagttcat agcccatata 300
tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 360
cccgccatt gagtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 420
attgacgtca atgggtggag tatttacggg aaactgcccc cttggcagta catcaagtgt 480
atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 540
atggccagta catgacctta tgggactttc ctacttggca gtacatctac gtattagtca 600
tcgctattac catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcgggtttg 660
actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc 720
aaaatcaacg ggaactttcca aaatgtcgta acaactccgc ccattgacg caaatgggcg 780
gtaggcgtgt acgggtgggag gtctatatata gcagagctct ctgggctaact agagaaccca 840
ctgcttactg gcttatcgaa attaatcaga ctacatatag ggagacccaa gcttggtaacc 900
gagctcggat cgatatctgc ggccgcgtcg acggaattca gtggatccac tagtaaccggc 960
cgccagtgtg ctggaattaa ttcgctgtct cgcaggggcca gctgttgggg tgagtactcc 1020
ctctcaaaag cgggcatgac ttctgcgcta agattgtcag tttccaaaaa cgaggaggat 1080
ttgatattca cctggcccgc ggtgatgcct ttgaggtggc ccgcgtccat ctggtcagaa 1140
aagacaatct ttttgttgtc aagcttgagg tctggcaggg ttgagatctg gccatacaact 1200
tgagtgatcga tgacatccac ttgaccttcc tctccacagg tgtccactcc caggtccaac 1260
tgaggtcga gcatgcatct agggcgggcca attccgcccc tctccctccc cccccctaa 1320

cgttactggc	cgaagccgct	tggaataagg	ccggtgtgag	tttgtctata	tgtgattttc	1380
caccatattg	ccgtctttttg	gcaatgtgag	ggcccggaag	cctggccctg	tcttcttgac	1440
gagcattcct	aggggtcttt	cccctctcgc	caaaggaaatg	caaggctctgt	tgaatgtcgt	1500
gaaggaagca	gttccctctgg	aagcttcttg	aagacaaaaca	acgtctgtag	cgaccctttg	1560
caggcagcgg	aacccccccac	ctggcgacag	gtgcctctgc	ggccaaaagc	cacgtgtata	1620
agatacacct	gcaaaggcgg	cacaacccca	gtgccacgtt	gtgagttgga	tagttgtgga	1680
aagagtcaaa	tggtctctcct	caagcgtatt	caacaagggg	ctgaaggatg	cccagaagg	1740
acccccattgt	atgggatctg	atctggggcc	tcggtgcaca	tgctttacat	gtgttttagtc	1800
gagggtaaaa	aaacgtctag	gccccccgaa	ccacgggggac	gtggtttttcc	tttgaaaaaac	1860
acgatgataa	gcttgccaca	acccgggatc	caccggtcgc	caccatgggtg	agcaagggcg	1920
aggagctggt	caccgggggtg	gtgcccattc	tggtcgagct	ggacggcgac	gtaaacggcc	1980
acaagttcag	cgtgtccggc	gagggcgagg	gcgatgccac	ctacggcaag	ctgaccctga	2040
agttcatctg	caccaccggc	aagctgcccc	tgccctggcc	caccctcgtg	accaccctga	2100
cctacggcgt	gcagtgcctt	agccgctacc	ccgaccacat	gaagcagcac	gacttcttga	2160
agtccgccat	gcccgaaggc	tacgtccagg	agcgaccat	cttcttcaag	gacgacggca	2220
actacaagac	ccgcgcgcgag	gtgaagtctg	agggcgacac	cctgggtgaac	cgcctcgagc	2280
tgaagggcac	cgacttcaag	gaggacggca	acatcctggg	gcacaagctg	gagtagaact	2340
acaacagcca	caacgtctat	atcatggccg	acaagcagaa	gaacggcatc	aagggtgaact	2400
tcaagatccg	ccacaacatc	gaggacggca	gcgtgcagct	cgccgaccac	taccagcaga	2460
acacccccat	cggcgcgcgc	cccgtgctgc	tgcccgcaca	ccactacctg	agcaccctga	2520
ccgccccgag	caaagaccac	aacgagaagc	gcgatcacat	ggtcctgctg	gagttcgtga	2580
ccgcccgcgg	gatcactctc	ggcatggacg	agctgtacaa	gtaaaagcggc	cctagagctc	2640
gctgatcagc	ctcgactgtg	cctctagtgt	ccagccatct	ggtgtttgccc	cctccccctg	2700
gccttccttg	accctgggaag	gtgccactcc	cactgtcctt	tcctaataaaa	atgaggaaat	2760
tgcatcgcat	tgtctgagta	ggtgtcattc	tattctgggg	ggtgggggtgg	ggcaggacag	2820
caagggggag	gattgggaag	acaatagcag	gcgatgctggg	gatgcgggtgg	gctctatggc	2880
ttctgaggcg	gaaagaacca	gctgggggct	gagtgcatct	tagttgtgggt	ttgtccaaac	2940
tcatcaatgt	atcttatcat	gtctgtatac	cgtcgacctc	tagctagagc	ttggcgtaat	3000
catgggtcata	gctgttttct	gtgtgaaatt	gttatccgct	cacaattcca	cacaacatac	3060
gagccggaag	cataaagtgt	aaagcctggg	gtgcctaattg	agttagctaa	ctcacattaa	3120
ttgctgttgc	ctcactgccc	gctttccagt	cgggaaacct	gtcgtgccag	ctgcattaat	3180
gaatcggcca	acgcgcgggg	agaggcggtt	tgctatttgg	gcgctcttcc	gcttctctgc	3240
tactgagcgc	gctgcgctcg	gtcgttcggc	tgccgcgagc	ggatatcagct	cactcaaaag	3300
cggtataacg	gttatccaca	gaatcagggg	ataacgcagg	aaagaacatg	tgagcaaaaag	3360
ggcagcaaaa	ggccaggaac	cgtaaaaagg	ccgcgttgct	ggcgtttttc	cataggctcc	3420
gccccctga	cgagcatcac	aaaaatcgac	gtcgaagtca	gaggtggcga	aaccgcacag	3480
gactataaag	ataccaggcg	tttccccctg	gaagctccct	cgtgcgctct	cctgttccga	3540
ccctgcccgt	taccggatac	ctgtccgcct	ttctcccttc	gggaagcgtg	gcgctttctc	3600
aatgctcagc	ctgtaggtat	ctcagttcgg	tgctcctcgt	tcgctccaag	ctgggctgtg	3660
tgacgaacc	ccccgttcag	cccgcaccgt	gcgccttata	cggttaactat	cgtcttgagt	3720
ccaacccggg	aagacacgac	ttatcgccac	tgccagcagc	cactggtaac	aggattagca	3780
gagcgaggta	tgtaggcggg	gctacagagt	tcttgaagtg	gtggcctaac	tacggctaca	3840
ctagaaggac	agtattttgg	atctgcgctc	tgctgaagcc	agttaccttc	ggaaaaagag	3900
ttggtagctc	ttgatccggc	aaacaaaacca	ccgctggtag	cggtgggttt	tttgtttgca	3960
tatatgagta	tacgcgcaga	aaaaaaggat	ctcaagaaga	tcctttgatc	ttttctacgg	4020
gggtctgacg	tcagtggaac	gaaaaactcac	gttaagggat	tttgggtcatg	agattatcaa	4080
aaaggatctt	cacctagatc	cttttaaat	aaaaatgaag	ttttaaatca	atctaaagta	4140
tatatgagta	aacttgggtc	gacagttacc	aatgcttaat	cagtgaaggca	cctatctcag	4200
cgatctgtct	atctcggttc	tccatagtgt	cctgactccc	cgtcgtgtag	ataactacga	4260
tacgggaggg	cttaccatct	ggccccagtg	ctgcaatgat	accgcgagac	ccacgcctcac	4320
cgggtccaga	tttatcagca	ataaacacag	cagccggaag	ggccgagcgc	agaagtggct	4380
ctgcaacttt	atccgcctcc	atccagtcta	ttaatgtgtg	ccgggaagct	agagtaagta	4440
gttcgcccag	taatatgttt	cgcaacgctg	ttgccattgc	tacaggcatc	gtgggtgtcac	4500
gctcctcggt	tggtatgggt	tcattcagct	ccggttccca	acgatcaagg	cgagttacat	4560
gatcccccat	gttgtgcaaaa	aaagcgggta	gctccttcgg	tcctccgatc	gttgtcagaa	4620
gtaagtgggc	cgcagtggtta	tcactcatgg	ttatggcagc	actgcataat	tctcttactg	4680
tcatgccatc	cgtaaagtgc	ttttctgtga	ctgggtgagta	ctcaaccaag	tcattctgag	4740
aatagtgtat	gcggcgaccg	agttgtctct	gcccggcgct	aatacgggat	aataccgcgc	4800
cacatagcag	aactttaaaa	gtgctcatca	ttggaaaaacg	ttcttcgggg	cgaaaaactct	4860
caaggtactt	accgctgttg	agatccagtt	cgatgtaacc	cactcgtgca	cccaactgat	4920
cttcagcatc	ttttactttc	accagcgttt	ctgggtgagc	aaaaaacagg	aggcaaaatg	4980
ccgcaaaaaa	gggaataagg	gcgacacgga	aatgttgaat	actcatactc	ttcctttttc	5040
aatattattg	cagggtttat	gaggtttacc	gtctcatgag	cggatacata	tttgaatgta	5100
tttagaaaaa	taaacaaaata	gggggttccgc	gcacatttcc	ccgaaaagtg	ccacctgacg	5160
tc						5162

<211> 5627
<212> DNA
<213> Artificial Sequence

<220>
<223> pMG plasmid from InvivoGen; IRES sequence modified
EMCV nucleotides 2736-3308

<400> 27
caccggcgaa ggaggcctag atctatcgat tgtacagcta gctcgacatg ataagataca 60
ttgatgagtt tggacaaacc acaactagaa tgcagtga aaatgcttt atttgtgaaa 120
tttgtgatgc tattgcttta tttgtgaaat ttgtgatgct attgctttat ttgtaaccaat 180
tataagctgc aataaacaag ttaacaacaa caattgcatt catttttatgt ttcagggttca 240
gggggaggtg tgggagggtt tttaaagcaa gtaaaacctc tacaaatgtg gtagatccat 300
ttaaatgtta attaagaaca tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa 360
ggccgcgttg ctggcggttt tccataggct ccgccccctc gacgagcatc acaaaaaatcg 420
acgctcaagt cagagggtggc gaaacccgac aggactataa agataccagg cgtttcccc 480
tggaagctcc ctctgtgcgt ctectgttcc gaccctgccg ctaccggat acctgtccgc 540
ctttctccct tcgggaagcg tggcgctttc tcatagctca cgctgtagggt atctcagttc 600
ggtgtaggtc gtctgcctca agctgggctg tgtgcacgaa cccccggtc agcccgaccg 660
ctgcgcctta tccggttaact atcgtcttga gtccaacccg gtaagacacg acttatcgcc 720
actggcagca gccactgggt acaggattag cagagcgagg tatgtaggcg gtgctacaga 780
gttcttgaag tgggtggccta actacggcta cactagaaga acagtatttg gtatctgcgc 840
tctgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaacaaac 900
caccgctggg agcgggtgggt tttttgtttg caagcagcag attacgcgca gaaaaaaagg 960
atctcaagaa gatcctttga tcttttctac ggggtctgac gctcagtggg acgaaaactc 1020
acgttaaggg atttttggtca tggctagtta attaagctgc aataaacaat cattattttc 1080
attggatctg tgtgttgggt ttttgtgtgg gcttggggga gggggaggcc agaattgactc 1140
caagagctac aggaaggcag gtcagagacc ccactggaca aacagtggct ggactctgca 1200
ccataacaca caatcaacag gggagttagc tggatcgagc tagagtccgt tacataactt 1260
acggtaaatg gccgcctgg ctgaccgccc aacgaccccc gccatttgac gtcaataatg 1320
acgtatgttc ccataagta gccaataggg actttccatt gacgtcaatg ggtggagtat 1380
ttacggtaaa ctgcccactt ggcagtagat caagtgtatc atatgccaag tacgccccct 1440
attgacgtca atgacggtaa atggccccgc tggcattatg cccagtacat gaccttatgg 1500
gactttccta ctgtggcagta catctacgta ttagtctcgc ctattaccat ggtgatgcgg 1560
ttttggcagt acatcaatgg gcgtggatag cggtttgact caccgggatt tccaagctc 1620
caccoccatg acgtcaatgg gagtttgttt tggcaccaaa atcaacggga ctttccaaaa 1680
tgtcgtaaca actccgcccc attgacgcaa atgggcggtg ggcgtgtacg gtgggagggtc 1740
tatataagca gagctcggtt agtgaaccgt cagatcgccct ggagacgcca tccacgctgt 1800
tttgacctcc atagaagaca ccgggaccga tccagcctcc gcggccggga acggtgcatt 1860
ggaacgcgga ttccccgtgc caagagttag gtaagtaccg cctatagagt ctataggccc 1920
acccccctgg ctcttatgc atgctatact gtttttggct tggggctctat acacccccgc 1980
ttcctcatgt tatagggtgat ggtatagctt agcctatagg tgtgggttat tgaccattat 2040
tgaccactcc cctattgggt acgatacttt ccattactaa tccataacat ggctctttgc 2100
cacaactctc tttattggct atatgccaat acactgtcct cagagactg acacgactc 2160
tgtattttta caggatgggg totcatttat tatttcaaaa ttccacatata caacaccacc 2220
gtccccagtg ccgcagttt ttattaaaca taacgtggga tctccacgcg aatctcgggt 2280
acgtgttccg gacatggggt ctctccgggt agcggcggag cttctacatc cgagccctgc 2340
tccatgcct ccagcgactc atggtcgctc ggcagctcct tgctcctaac agtggaggcc 2400
agacttaggc acagcacgat gccaccacc accagtgtgc cgcacaaggc cgtggcggtg 2460
gggtatgtgt ctgaaaatga gctcggggag cgggcttgca ccgctgacgc atttggaaag 2520
cttaaggcag cggcagaaga agatgcaggc agctgagttg ttgtgttctg ataagagtca 2580
gaggtaactc ccgttgcggt gctgttaacg gtggagggca gtgtagtctg agcagtaact 2640
gttgctgccg cgcgcgccac cagacataat agctgacaga ctaacagact gttcctttcc 2700
atgggtcttt tctgcagtca ccggggggat ccttcgaacg tagctctaga ttgagtcgac 2760
gttactggcc gaagccgctt ggaataaggc cgggtgtcgt ttgtctatat gttattttcc 2820
accatattgc cgtcttttgg caatgtgagg gcccggaac ctggccctgt cttcttgacg 2880
agcattccta ggggtctttc cctctcgcg aaagggaatg aaggtctgtt gaatgtcgtg 2940
aaggaagcag ttccctctga agcttcttga agacaaacaa cgtctgtagc gaccttttgc 3000
aggcagcgga accccccacc tggcgacagg tggcctctgc gcaaaaagcc acgtgtataa 3060
gatacacctg caaaggcggc acaaccccag tgccacgttg tgagttggat agttgtggaa 3120
agagtcacaa ggctctctc aagcgtattc aacaaggggc tgaaggatgc ccagaaggta 3180
ccccattgta tggggtcct tctggggcct cgggtcacat gctttacatg tgttttagtc 3240
aggttaaaaa aacgtctagg cccccgaac cacggggacg tggttttcct ttgaaaaaca 3300
cgataatacc atgggtaagt gatattctat agttgtgacc ggcgcctagt gttgacaatt 3360
aatcatcggc atagatatc ggcatagtat actataggag ggccaccatg 3420
tcgactacta accttcttct ctttcttaca gctgagatca ccggtaggag ggccatcatg 3480

```

aaaaagcctg aactcaccgc gacgtctgtc gcgaagtttc tgatcgaaaa gttcgcacagc 3540
gtctccgacc tgatgcagct ctccggaggc gaagaatctc gtgctttcag cttcgatgta 3600
ggagggcgctg gatattgtcct gccgggtaaat agctgcgcgc atgggtttcta caaagatcgt 3660
tatgtttatc ggcactttgc atcgcccgcg ctcccgattc cggaaagtgt tgacattggg 3720
gaattcagcg agagcctgac ctattgcac tcccgccgtg cacagggtgt cacgttgcaa 3780
gacctgcctg aaaccgaact gcccgcgtgt ctgcaacccg tcgcggagct catggatgcg 3840
atcgctgcgg ccgatcttag ccagacgagc ggggttcggc cattcggacc gcaaggaaac 3900
ggccaataca ctacatggcg tgatttcata tgcgcgattg ctgatcccca tgtgtatcac 3960
tggcaaaactg tgatggacga caccgtcagt gcgtccgtcg cgcaggctct cgatgagctg 4020
atgctttggg ccgaggactg ccccgaagtc cggcacctcg tgcacgcgga tttcggctcc 4080
aacaatgtcc tgacggacaa tggccgcata acagcgggtc ttgactggag cgaggcgatg 4140
ttcgggggatt cccaatacga ggtcgccaac atcttcttct ggaggccgtg gttggcttgt 4200
atggagcagc agacgcgcta cttcgagcgg aggcacctcg agcttgagg atcgcccgcg 4260
ctccgggcgt atatgctccg cattggtctt gaccaactct atcagagctt ggttgacggc 4320
aatttcgatg atgcagcttg ggcgcagggt cgatgcgcag caatcgtccg atccggagcc 4380
gggactgtcg ggcgtacaca aatcgccccg agaagcgcg ccgtctggac cgatggctgt 4440
gtagaagtac tcgccgatag tggaaaaccga cgccccagca ctgctccgag ggcaaaggaa 4500
tgagtcgaga attcgctaga gggccctatt ctatagtgtc acctaaatgc tagagctcgc 4560
tgatcagcct cgactgtgcc ttctagttag cagccatctg ttgtttgccc ctcccctgtg 4620
ccttccttga ccctgggaagg tgccactccc actgtccttt cctaataaaa tgaggaaatt 4680
gcatcgcatt gtctgagtag gtgtcattct attctggggg gtgggggtggg gcaggacagc 4740
aagggggagg attgggaaga caatagcagg catgctcagg gcccaattgc tcgagcggcc 4800
gcaataaaaat atctttatct tcattacatc tgtgtgttgg ttttttgtgt gaatcgtaac 4860
taacatacgc tctccatcaa aacaaaacga aacaaaacaa actagcaaaa taggctgtcc 4920
ccagtgcag tgccaggtgcc agaactttc tctatcgaag gatctgcgat cgctccgggtg 4980
cccgtcagtg ggcagagcgc acatcgccca cagtcgccga gaagttgggg ggaggggtcg 5040
gcaattgaac cgggtgcctag agaaggtggc gcggggtaaa ctgggaaagt gatgtcgtgt 5100
actggctccg cctttttccc gaggttgggg gagaaccgta tataagtga gtagtcgccg 5160
tgaacgttct ttttcgcaac gggtttgccg ccagaacaca gctgaagctt cgaggggtcg 5220
gcatctctcc ttcacgcgcc cgccgcccta cctgaggccg ccacccacgc cgttgagtc 5280
gcgttctgcc gccctccgcc tgtggtgccc cctgaactgc gtccgcgcgc taggtaagtt 5340
taaagctcag ctcgagaccg ggcctttgtc cggcgctccc ttggagccta cctagactca 5400
gccggtcttc cgcgttttgc ctgacctgc ttgctcaact ctacgtcttt gtttcgtttt 5460
ctgttctgcg ccgttacaga tccaagctgt gaccggcgcc tacgtaagtg atatatcta 5520
gattttacaa aaagagtgtt gacttgtgag cgctcacaa tgcatactag attcatcgag 5580
agggacacgt cgactactaa ccttcttctc tttcctacag ctgagat 5627

```

<210> 28
 <211> 553
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pMG plasmid from InvivoGen: EMCV IRES sequence

```

<400> 28
aacgttactg gccgaagcgc cttggaataa ggccggtgtg cgtttgtcta tatgttattt 60
tccaccatat tgccgtcttt tggcaatgtg agggcccgga aacctggccc tgtcttcttg 120
acgagcattc ctaggggtct tccccctctc gccaaaggaa tgcaaggctc gttgaatgtc 180
gtgaaggaag cagttcctct ggaagcttct aggtgcctct gcgccaaaaa gccacgtgta 240
tgcaggcagc ggaacccccc acctggcgac cagtgccacg ttgtgagttg gatagttgtg 300
taagatacac ctgcaaaggc ggcacaaccc cagtggcagc ttgtgagttg gatagttgtg 360
gaaagagtca aatggctctc ctcaagcgta ttcaacaagg ggctgaagga tgcacagaag 420
gtaccccatg gtatgggatc tgatctgggg cctcggtgca catgctttac gtgtgtttag 480
tcgaggttaa aaaacgtcta gggcccccca accacgggga cgtgggtttc ctttgaaaaa 540
cacgatgata ata 553

```

<210> 29
 <211> 4692
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pDSred1-N1 plasmid from Clontech

```

<400> 29
tagttattaa tagtaatcaa ttacgggggtc attagttcat agcccatata tggagttccg 60

```

cggtacataa	cttacggtaa	atggccccgc	tggttgaccg	cccaacgacc	cccgccatt	120
gacgtcaata	atgacgtatg	ttcccatagt	aacgcccaata	gggactttcc	attgacgtca	180
atgggtggag	tattttacgg	aaactgcccc	cttggcagta	catcaagtgt	atcatatgcc	240
aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgcccagta	300
catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	tcgctattac	360
catggtgatg	cggttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	420
atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	480
ggactttcca	aaatgtcgta	acaactccgc	cccatgacg	caaatgggcg	gtaggcgtgt	540
acgggtgggag	gtctatataa	gcagagctgg	tttagtgaac	cgtcagatcc	gctagcgcta	600
ccggactcag	atctcgagct	caagcttcga	attctgcagt	cgacggtaac	cggggcccgg	660
gatccaccgg	tgcgccaccat	ggtgcgctcc	tccaagaacg	tcatcaagga	gttcatgcgc	720
ttcaaggtgc	gcatggagggg	caccgtgaac	ggccacgagt	tcgagatcga	gggcgagggc	780
gagggccgcc	cctacgaggg	ccacaacacc	gtgaagctga	aggtgaccaa	ggcgggcccc	840
ctgcccctcg	cctgggacat	cctgtccccc	cagttccagt	acggctccaa	ggtgtacgtg	900
aagcaccccc	ccgacatccc	cgactacaag	aagctgtcct	tccccgaggg	cttcaagttg	960
gagcgcgtga	tgaacttcga	ggacggcggc	gtggtgaccg	tgacccagga	ctccctctgt	1020
caggacgggt	gcttcatcta	caaggtgaag	ttcatcgcg	tgaacttccc	ctccgacggc	1080
cccgtaatgc	agaagaagac	catgggctgg	gagggcctca	ccgagcgcct	gtacccccgc	1140
gacggcgtgc	tgaagggcga	gatccacaag	gccttgagc	tgaaggacgg	cggccactac	1200
ctggtggagt	tcaagttccat	ctacatggcc	aagaagcccg	tgacgtgcc	cggctactac	1260
tacgtggact	ccaagctgga	catcacctcc	cacaacgagg	actacaccat	cgtggagcag	1320
tacgagcgca	ccgagggccg	ccaccactcg	ttcctgtagc	ggccgcgact	ctagatcata	1380
atcagccata	ccacatttgt	agaggtttta	cttgctttaa	aaaaacctccc	acacctcccc	1440
ctgaacctga	aacataaaat	gaatgcaatt	gttggtgtta	acttgtttat	tcgagcttat	1500
aatggttaca	aataaagcaa	tagcatcaca	aatttcacaa	ataaagcatt	tttttactg	1560
cattctagtt	gtggtttgtc	caaaactcatc	aatgtatctt	aaggcgtaaa	ttgtaagcgt	1620
taatattttg	ttaaaattcg	cgttaaaatt	ttgttaaatc	agctcatttt	ttaaccaata	1680
ggccgaaatc	ggcaaaatcc	cttataaatc	aaaagaatag	accgagatag	ggttgagtgt	1740
ttgtccagtt	tggaacaaga	gtccactatt	aaagaacgtg	gactccaacg	tcaagggcgc	1800
aaaaaccgtc	tatcagggcg	atggcccact	acgtgaacca	tcaccctaata	caagtttttt	1860
ggggtcgagg	tgccgtaaag	cactaaatcg	gaaccctaaa	gggagccccc	gatttagagc	1920
ttgacgggga	aagccggcga	acgtggcgag	aaaggaaggg	aagaaagcga	aaggagcggg	1980
cgctagggcg	ctggcaagtg	tagcggtcac	gctgcgcgta	accaccacac	ccgcccgcgt	2040
taatgcgcgc	ctacagggcg	cgtcaggtgg	cacttttcgg	ggaaatgtgc	gcggaacccc	2100
tatttgttta	tttttctaaa	tacattcaaa	tatgtatccg	ctcatgagac	aataaccctg	2160
ataaatgctt	caataatatt	gaaaaaggaa	gagtcctgag	gcggaaagaa	ccagctgtgg	2220
aatgtgtgtc	agttaggggtg	tggaaagtcc	ccaggctccc	cagcaggcag	aagtatgcaa	2280
agcatgcata	tcaattagtc	agcaaccagg	tgtggaaaagt	ccccaggctc	cccagcggc	2340
agaagtatgc	aaagcatgca	tctcaattag	tcagcaacca	tagtcccgcc	cctaactccg	2400
cccatccgcg	ccctaactcc	gcccagttcc	gcccattctc	cgccccatgg	ctgactaatt	2460
ttttttattt	atgcagaggc	cgaggccgcc	ctggcctctg	agctattcca	gaagtgtga	2520
ggaggctttt	ttggaggcct	aggcttttgc	aaagatcgat	caagagacag	gatgaggatc	2580
gttttcgcatg	attgaacaag	atggattgca	cgcaggttct	ccggccgctt	gggtggagag	2640
gctattcggc	tatgactggg	cacaacagac	aatcggctgc	cttgatgccg	ccgtgttccg	2700
gctgtcagcg	cagggggcgc	cggttctttt	tgtcaagacc	gacctgtccg	gtgccctgaa	2760
tgaactgcaa	gacgaggcag	cgcggtatc	gtggctggcc	acgacgggcg	ttccttgcgc	2820
agctgtgtc	gacgttgtca	ctgaagcggg	aagggactgg	ctgctattgg	gcgaagtgcc	2880
ggggcaggat	ctcctgtcat	ctcaccttgc	tcctgccgag	aaagtatcca	tcatggctga	2940
tgcaatgcgg	cggtgtcata	cgcttgatcc	ggctacctgc	ccattcgacc	accaagcgaa	3000
acatcgcata	gagcagcac	gtactcggat	ggaagccggg	cttgtcgatc	aggatgatct	3060
ggacgaagag	catcaggggc	tcgcgccagc	cgaactgttc	gccaggctca	aggcagcat	3120
gcccagcggc	gaggatctcg	tcgtgaccga	tggcgatgcc	tgcttgccga	atatcatggt	3180
ggaaaatggc	cgcttttctg	gattcatcga	ctgtggccgg	ctgggtgtgg	cggaccgcta	3240
tcaggacata	gcgttggcta	cccgtgatat	tgctgaagag	cttggcggcg	aatgggctga	3300
ccgcttccct	gtgctttacg	gtatcgcgcg	tcccgatccg	cagcgcacat	ccttctatcg	3360
ccttcttgac	gagctcttct	gagcgggact	ctgggggttcg	aaatgaccga	ccaagcgacg	3420
cccaacctgc	catcacgaga	tttcgattcc	accggccgct	tctatgaaag	gttgggcttc	3480
ggaatcgttt	tcggggacgc	cggtctggatg	atcctccagc	gcgggggatct	catgctggag	3540
ttcttcgccc	accctagggg	gaggctaact	gaaacacgga	aggagacaat	accggaagga	3600
acccgcgcta	tgacggcaat	aaaaagacag	aataaaacgc	acgggtgttg	gtcgttgtgt	3660
cataaacgcg	gggttcgggtc	ccagggctgg	cactctgtgc	atacccccacc	gagaccccat	3720
tggggccaat	acgcccgcgt	ttcttccctt	tccccacccc	accccccaag	ttcgggtgaa	3780
ggccaggggc	tcgcagccaa	cgctcggggcg	gcaggccctg	ccatagcctc	aggttactca	3840
tatatacttt	agattgattt	aaaacttcat	ttttaattta	aaaggatcta	ggtgaagatc	3900
ctttttgata	atctcatgac	caaaatccct	taacgtgagt	tttcgttcca	ctgagcgtca	3960
gaccccgtag	aaaagatcaa	aggatcttct	tgagatccct	tttttctgcg	cgtaaatctgc	4020
tgcttgcaaa	caaaaaaacc	accgctacca	gcggtggttt	gtttgccgga	tcaagagcta	4080

ccaactcttt	ttccgaaggt	aactggcttc	agcagagcgc	agataccaaa	tactgtcctt	4140
ctagtgtagc	cgtagttagg	ccaccacttc	aagaactctg	tagcaccgcc	tacatacctc	4200
gctctgtctaa	tcctgtttacc	agtggctgct	gccagtgccg	ataagtctgt	tcttaccggg	4260
ttggactcaa	gacgatagtt	accggataag	gcgcagcggg	cgggctgaac	gggggggttcg	4320
tgcacacagc	ccagcttggg	gcgaacgacc	tacaccgaac	tgagatacct	acagcgtgag	4380
ctatgagaaa	gcgccacgct	tcccgaaggg	agaaaggcgg	acaggtatcc	ggtaagcggc	4440
agggtcggaa	caggagagcg	cacgagggag	cttccagggg	gaaacgcctg	gtatctttat	4500
agtcctgtcg	ggtttcgcca	cctctgactt	gagcgtcgat	ttttgtgatg	ctcgtcaggg	4560
gggcggagcc	tatggaaaaa	cgccagcaac	gcggcctttt	tacggttcct	ggccttttgc	4620
tggccttttg	ctcacatgtt	ctttcctgcg	ttatcccctg	attctgtgga	taaccgtatt	4680
accgccatgc	at					4692

<210> 30

<211> 4257

<212> DNA

<213> Artificial Sequence

<220>

<223> pPur plasmid from Clontech

<400> 30

ctgtggaatg	tgtgtcagtt	aggggtgtgga	aagtccccag	gctccccagc	aggcagaagt	60
atgcaaagca	tgcattctcaa	ttagtcagca	accaggtgtg	gaaagtcccc	aggctcccca	120
gcaggcagaa	gtatgcaaag	catgcatctc	aattagtcag	caaccatagt	cccgccccta	180
actccgcccc	tcccggccct	aactccgccc	agttccgccc	attctccgcc	ccatggctga	240
ctaatttttt	ttattttatgc	agaggccgag	gccgcctcgg	cctctgagct	attccagaag	300
tagtgaggag	gcttttttgg	aggccttaggc	ttttgcaaaa	agcttgcatt	cctgcaggtc	360
ggccgcccacg	accggtgccc	ccaccatccc	ctgacccacg	ccctgacccc	ctcacaagga	420
gacgaaccttc	catgaccgag	tacaagccca	cgggtgcgct	cgccaccgcg	gacgacgtcc	480
cccggggccgt	acgcaccctc	gccgcccgcg	tcgcccacta	ccccgccacg	cgccacaccg	540
tgcaccccgga	ccgccacatc	gagcgggtca	ccgagctgca	agaactcttc	ctcacgcgcg	600
tcgggctcga	catcggcaag	gtgtgggtcg	cggacgacgg	cgccgcgggtg	gcggctctgga	660
ccacgcggga	gagcgtcgaa	gcggggggcgg	tggttcgcca	gategggccc	cgcatggccc	720
agttgagcgg	ttcccggctg	gccgcgcagc	aacagatgga	aggcctcctg	gcgcgcgacc	780
ggcccaagga	gcccgcgctg	ttcctggcca	ccgtcggcgt	ctcgcccagc	caccagggga	840
aggggtctggg	cagcgccgtc	gtgctccccg	gagtgaggcg	ggccgagcgc	gcccgggtgc	900
ccgccttcct	ggagacctcc	gcgccccgca	acctcccctt	ctacgagcgg	ctcggcttca	960
ccgtcaccgc	cgagctcgag	gtgcccgaag	gaccgcgcac	ctggtgcatg	accgcgaagc	1020
ccgtgacctg	acgcgccgcc	cacgaccccg	agcgcgccag	cgaaaggagc	gcacgacccc	1080
atgggtccga	ccgaagccga	cccggggcggc	ccgcgccgac	ccgcacccgc	ccccgaggcc	1140
caccgactct	agaggatcat	aatcagccat	accacatttg	tagaggtttt	acttgcttta	1200
aaaaacctcc	cacacctccc	cctgaacctg	aaacataaaa	tgaatgcaat	tggtgttggt	1260
aacttggtta	ttgcagctta	taatgggtac	aaataaagca	atagcatcac	aaatttcaca	1320
aataaagcat	ttttttcaat	gcattctagt	tggtggtttg	ccaaactcat	caatgtatct	1380
tatcatgtct	ggatccccag	gaagctcctc	tggtgtcctc	ttaaccctaa	cctcctctac	1440
ttgagaggac	attccaatca	taggctgccc	atccaccctc	tggtgtcctc	tggttaattag	1500
gtcacttaac	aaaaaggaaa	ttgggtaggg	gtttttcaca	gaccgctttc	taagggtaat	1560
tttaaaatat	ctgggaagtc	ccttccactg	ctgtgttcca	gaagtgttgg	taaacagccc	1620
acaaatgtca	acagcagaaa	catacaagct	gtcagctttg	cacaagggcc	caacaccctg	1680
ctcatcaaga	agcactgtgg	ttgctgtggt	agtaatgtgc	aaaacaggag	gcacattttc	1740
cccacctgtg	taggttccaa	aatatctagt	gttttccatt	ttacttggat	caggaaacca	1800
gcactccact	ggataagcat	tatccttata	caaaacagcc	ttgtgggtcag	tggtcatctg	1860
ctgactgtca	actgtagcat	tttttggggg	tacagtttga	gcaggatatt	tggtcctgta	1920
gtttgctaac	acaccctgca	gctccaaagg	ttccccacca	acagcaaaaa	aatgaaaatt	1980
tgaccttga	atgggttttc	cagcaccatt	ttcatgagtt	ttttgtgtcc	ctgaatgcaa	2040
gtttaacata	gcagttaccc	caataacctc	agttttaaca	gtaacagctt	cccacatcaa	2100
aatattttcca	caggttaagt	cctcatttaa	attaggcaaa	ggaattcttg	aagacgaaag	2160
ggcctcgtga	tacgcctatt	tttataggtt	aatgtcatga	taataatggg	ttcttagacg	2220
tcagggtggca	cttttcgggg	aaatgtgcgc	ggaaccccta	tttgtttatt	tttctaaata	2280
catttgggtg	tgtataccgt	catgagacaa	taacctgat	aaatgcttca	ataatattga	2340
aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttattccctt	ttttgcggca	2400
ttttgccttc	ctgtttttgc	tcacccagaa	acgctggtga	aagtaaaaga	tgctgaagat	2460
cagttgggtg	ccagagtggg	ttacatcgaa	ctggatctca	acagcggtaa	gatccttgag	2520
agttttcgcc	ccgaagaacg	ttttccaatg	atgagcactt	ttaaagtctt	gctatgtggc	2580
gcggtattat	ccggtgttga	cgccgggcaa	gagcaactcg	gtcgcgcgat	acactattct	2640
cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	tggaatgaca	2700
gtaagagaat	tatgcagtgc	tgccataacc	atgagtgtga	acactgcggc	caacttactt	2760

ctgacaacga	tccggaggacc	gaaggagcta	accgcttttt	tgcacaacat	gggggatcat	2820
gtaactcgcc	ttgatcggtt	ggaaccggag	ctgaatgaag	ccataccaaa	cgacgagcgt	2880
gacaccacga	tgcctgcagc	aatggcaaca	acgttgcgca	aactattaac	tggcgaacta	2940
cttactctag	cttcccggca	acaattaata	gactggatgg	aggcggataa	agttgcagga	3000
ccacttctgc	gctcggccct	tccggctggc	tggtttattg	ctgataaatc	tggagccggt	3060
gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	atggtaagcc	ctcccgtatc	3120
gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	acagatcgct	3180
gagatagggt	cctcactgat	taagcattgg	taactgtcag	accaagttta	ctcatatata	3240
cttttagattg	atttaaaact	tcatttttaa	tttaaaagga	tctaggtgaa	gatccttttt	3300
gataatctca	tgacaaaaat	cccttaacgt	gagttttcgt	tccactgagc	gtcagacccc	3360
gtagaaaaaga	tcaaaggatc	ttcttgagat	cccttttttc	tgcgcgtaat	ctgctgcttg	3420
caaacaaaaa	aaccaccgct	accagcggtg	gtttgtttgc	cggatcaaga	gctaccaact	3480
ctttttccga	aggtaactgg	cttcagcaga	gcgcagatac	caaatactgt	ccttctagt	3540
tagccgtagt	taggccaaca	cttcaagaac	tctgtagcac	cgctacata	cctcgctctg	3600
ctaatacctgt	taccagtggc	tgctgccagt	ggcgataagt	cgtgtcttac	cgggttggac	3660
tcaagacgat	agttaccgga	taaggcgag	cggtcggggt	gaacgggggg	ttcgtgcaca	3720
cagcccagct	tggagcgaa	gacctacacc	gaactgagat	acctacagcg	tgagctatga	3780
gaaagcgcca	cgcttcccga	agggagaaa	gcggacaggt	atccggtaag	cggcagggtc	3840
ggaacaggag	agcgcacgag	ggagcttcca	gggggaaaac	cctgggtatct	ttatagtcct	3900
gtcgggtttc	gccacctctg	acttgagcgt	cgatttttgc	gatgctcgtc	agggggcg	3960
agcctatgga	aaaacgccag	caacgcggcc	tttttacggg	tcctggcctt	ttgctggcct	4020
tttgctcaca	tggttcttcc	tgcttatccc	cctgattctg	tggataaccg	tattaccgcc	4080
tttgagttag	ctgataccgc	tcgcccagc	cgaacgacgg	agcgacgaga	gtcagtgagc	4140
gaggaagcgg	aagagcgct	gatgcgggat	tttctcctta	cgcatctgtg	cggattttca	4200
caccgcatat	ggtgcactct	cagtacaatc	tgctctgatg	ccgcatagtt	aagccag	4257

<210> 31
 <211> 8136
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pWE15 cosmid vector

<300>
 <308> GenBank X65279
 <309> 1995-04-14

<400> 31						
ctatagttag	tcgattattg	cggccgcgaa	ttcttgaaga	cgaaagggcc	tcgtgatacg	60
cctattttta	taggttaaat	tcatgataat	aatgggtttc	tagacgtcag	gtggcacttt	120
tccgggaaat	gtgcgcgga	cccctatttg	tttatttttc	taaatacatt	caaataatga	180
tccgctcatg	agacaataac	cctgataaat	gcttcaataa	tattgaaaaa	ggaagagtat	240
gagatttcaa	catttccgtg	tcgcccttat	tccttttttt	gcggcatttt	gcttccgtgt	300
tttgctcacc	cagaaaacgc	ggtgaaagta	aaagatgctg	aagatcagtt	gggtgcacga	360
gtgggttaca	tcgaactgga	tctcaacagc	ggtaagatcc	ttgagagttt	tcgccccgaa	420
gaacggtttc	caatgatgag	cactttttaa	gttctgctat	gtggcgcggt	attatcccgt	480
gttgacgccc	ggcaagagca	actcggctcg	cgcatacact	attctcagaa	tgacttgggt	540
gagtactcac	cagtcacaga	aaagcatctt	acggatggca	tgacagtaag	agaattatgc	600
agtgtcgcca	taacctagag	tgataaacat	gcggccaact	tacttctgac	aacgatcggg	660
ggaccgaagg	agctaaccgc	ttttttgcac	aacatggggg	atcatgtaac	tcgccttgat	720
cgttggggaa	cggagctgaa	tgaagccata	ccaaacgcag	agcgtgacac	cacgatgcct	780
gcagcaatgg	caacaacggt	gcgcaaaact	ttaactggcg	aactacttac	tctagcttcc	840
cggcaacaat	taatagactg	gatggaggcg	gataaaagtt	caggaccact	tctgcgctcg	900
gcccttccgg	ctggctgggt	tattgctgat	aaatctggag	ccggtgagcg	tgggtctcgc	960
ggtatcattg	cagcagctgg	gccagatggg	aagccctccc	gtatcgtagt	tatctacacg	1020
acggggagtc	aggcaactat	ggatgaacga	aatagacaga	tcgctgagat	aggtgcctca	1080
ctgattaagc	attggtaact	gtcagaccaa	gtttactcat	atatacttta	gattgattta	1140
aaacttcatt	tttaatttaa	aaggatctag	gtgaagatcc	tttttgataa	tctcatgacc	1200
aaaatccctt	aacgtgagtt	ttcgttccac	tgagcgtcag	accccgtaga	aaagatcaaa	1260
ggatcttctt	gagatccctt	ttttctgcgc	gtaatctgct	gcttgcaaac	aaaaaaacca	1320
ccgctaccag	cgggtggttt	tttgccggat	caagagctac	caactctttt	tccgaaggta	1380
actggcttca	gcagagcgca	gataccaaat	actgtccttc	tagtgtagcc	gtagtttagc	1440
caccacttca	agaactctgt	agcaccgcct	acatacctcg	ctctgctaata	cctgttacca	1500
gtggctgctg	ccagtggcga	taagtgcgtg	cttaccgggt	tggactcaag	acgatagtta	1560
cgggataaag	cgcagcggct	gggctgaacg	gggggttcgt	gcacacagcc	cagcttggag	1620
cgaacgacct	acaccgaact	gagataccta	cagcgtgagc	tatgagaaag	cgccacgctt	1680

ccgaagggag	aaaggcggac	aggatatccgg	taagcggcag	ggtcggaaca	ggagagcgca	1740
cgagggagct	tccaggggga	aacgcctggg	atctttatag	tcctgtcggy	gtttcgccac	1800
ctctgacttg	agcgctcgatt	tttgtgatgc	tcgtcagggg	ggcggagcct	atggaaaaac	1860
gccagcaacg	cggccttttt	acggttcctg	gccttttgct	ggccttttgc	tcacatgttc	1920
tttctgcgt	tatccctga	ttctgtggat	aaccgtatta	ccgcctttga	gtgagctgat	1980
accgctcgcc	gcagccgaac	gaccgagcgc	agcgagtcag	tgagcgagga	agcgggaagag	2040
cgctgacttc	cgcgtttcca	gactttacga	aacacggaaa	ccgaagacca	ttcatgttgt	2100
tgctcaggtc	gcagacgttt	tgacgacgca	gtcgcttcac	gttcgctcgc	gtatcgggtga	2160
ttcattctgc	taaccagtaa	ggcaaccccg	ccagcctagc	cgggtcctca	acgacaggag	2220
cacgatcatg	cgcacccgtc	agatccagac	atgataagat	acattgatga	gtttggacaa	2280
accacaacta	gaatgcagtg	aaaaaaatgc	tttattttgtg	aaattttgtga	tgctattgct	2340
ttattttgtaa	ccattataag	ctgcaataaaa	caagttaaca	acaacaattg	cattcatttt	2400
atgtttcagg	ttcaggggga	gggtgtgggag	gttttttaaa	gcaagtataaa	cctctacaaa	2460
tgtggtatgg	ctgattatga	tctctagtca	aggcactata	catcaaatat	tccttattaa	2520
cccttttaca	aattaaaaag	ctaaagggtac	acaattttttg	agcatagtta	ttaatagcag	2580
acactctatg	cctgtgtgga	gtaagaaaaa	acagtatgtt	atgattataa	ctgtttatgcc	2640
tactataaaa	gggttacagaa	tattttttcca	taattttctt	gtatagcagt	gcagcttttt	2700
cctttgtggt	gtaaatagca	aagcaagcaa	gagttctatt	actaaacaca	gcatgactca	2760
aaaaacttag	caattctgaa	ggaaagtctt	tgggtctctc	tacctttctc	ttcttttttg	2820
gaggagttag	atgttgagag	tcagcagtag	cctcatcatc	actagatggc	atttcttctg	2880
agcaaaacag	gttttcctca	ttaaaggcat	tccaccactg	ctcccattca	tcagttccat	2940
aggttggaat	ctaaaaatata	caaacaatta	gaatcagtag	tttaacacat	tatacactta	3000
aaaattttat	atttacctta	gagcttttaa	tctctgtagg	tagtttgtcc	aattatgtca	3060
caccacagaa	gtaagggtcc	ttcacaaaaga	tccggaccaaa	agcggccatc	gtgcctcccc	3120
actcctgcag	ttcgggggca	tggatgcgcg	gatagccgct	gctgggtttcc	tggtatgccga	3180
cggatttgca	ctgcgcgtag	aactcgcgag	gtcgtccagc	ctcaggcagc	agctgaacca	3240
actcgcgagg	ggatcgagcc	cgggggtgggc	gaagaactcc	agcatgagat	ccccgcgctg	3300
gaggatcatc	cagccggcgt	cccggaaaaac	gattccgaag	cccaaccttt	catagaaggc	3360
ggcgggtgaa	tcgaaatctc	gtgatggcag	gttgggcgtc	gcttggtcgg	tcatttcgaa	3420
ccccagagtc	ccgctcagaa	gaactcgtca	agaaggcgat	agaaggcgat	gcgctgcgaa	3480
tcgggagcgg	cgataccgta	aagcacgagg	aagcggtcag	cccattcgcc	gccaagctct	3540
tcagcaatat	caacggtagc	caacgctatg	tccgtgatgc	ggtcgcgccac	accagccgg	3600
ccacagtcga	tgaatccaga	aaagcggcca	ttttccacca	tgatattcgg	caagcaggca	3660
tcgccatggg	tcacgacgag	atcctcgccc	tccgggatgc	cgccttgagc	ctggcgaaaca	3720
gttcggctgg	cgcgcggccc	tgatgctctt	cgctcgatgc	atcctgatcg	acaagaccgg	3780
cttccatccg	agtagctgct	cgctcgatgc	agccgcgcga	ttgggtggtcg	aatgggcagg	3840
tagccggatc	aagcgtatgc	agccgcgcga	ttgcatcagc	catgatggat	actttctcgg	3900
caggagatgc	gtgagatgac	aggagatcct	gccccggcac	ttcgcccaat	agcagccagt	3960
cccttcccgc	ttcagtgaca	acgtcgagca	cagctgcgca	aggaacgccc	gtcgtggcca	4020
gccacgatag	ccgcgctgcc	tcgtcctgca	gttcattcag	ggcaccggac	aggtcggtct	4080
tgacaaaaag	aaccgggcgc	ccctgcgctg	acagccggaa	cacggcggga	tcagagccagc	4140
cgattgtctg	ttgtgcccag	tcatagccga	atagcctctc	caccgaagcg	gccggagaac	4200
ctgcgtgcaa	tccatcttgt	tcaatcatgc	gaaacgatcc	tcattcctgtc	tcttgatcag	4260
atcttgatcc	cctgcgccat	cagatccttg	gcggcaagaa	agccatccag	tttactttgc	4320
agggcttccc	aaccttacca	gagggcgccc	cagctggcaa	ttccgggttcg	cttgcctgtcc	4380
ataaaaccgc	ccagtcctagc	tatcgccatg	taagcccaat	gcaagctacc	tgctttctct	4440
ttgcgcttgc	gttttccctt	gtccagatag	cccagtagct	gacattccatc	cggggtcagc	4500
accgtttctg	cggaactggct	ttctacgtgt	tccgcttctt	ttagcagccc	ttgcgcctctg	4560
agtgccttgc	gcagcgtgaa	agcttttttg	aaaagcctag	gcctccaaaa	aagcctcctc	4620
actacttctg	gaatagctca	gagggcggag	cgcctaaat	aaaaaaaatt	agtcagccat	4680
ggggcgagga	atggggcgga	ctgggcggag	ttaggggcgg	gatgggcgga	gttaggggagc	4740
ggactatggt	tgctgactaa	ttgagatgca	tgctttgcct	acttctgcct	gctggggagc	4800
ctggggactt	tccacacctg	gttgctgact	aattgagatg	catgctttgc	atacttctgc	4860
ctgctgggga	gcctggggac	tttccacacc	ctaactgaca	cacattccac	agccggatct	4920
gcaggaccca	acgctgcocg	agatgcgcgc	cgtgcagctg	ctggagatgg	cggacgcgat	4980
ggatatgttc	tgccaagggt	tgggttgccg	attcagctgt	ctccgcaaga	attgattggc	5040
tccaattctt	ggagtgggtg	atccgttagc	gaggtgcgcg	cggcttccat	tcaggtcgag	5100
gtggcccggc	tccatgcacc	gcgacgcacc	gcggggaggc	agacaaggta	tagggcgggcg	5160
cctacaatcc	atgccaaccc	gttccatgtg	ctgcgcgagg	cgcataaatc	gccgtgacga	5220
tcagcggtcc	aatgatcgaa	gttaggctgg	taagacgcgc	gagcgatcct	tgaagctgtc	5280
cctgatgggt	gtcatctacc	tgccctggaca	gcattggcctg	caacgcggga	tcccgatgcc	5340
gccggaagcg	agaagaatca	taatggggaa	ggccatccag	cctcgcgtcg	cgaacgcgag	5400
caagacgtag	cccagcgctg	cgggcgcgca	tgccggcgat	aatggcctgc	ttctcgccga	5460
aacgtttggt	ggcggggacca	gtgacgaagg	cttgagcgag	ggcgtgcaag	attccgaata	5520
ccgcaagcga	caggccgatc	atcgtcgcgc	tccagcgaaa	gcggctcctcg	ccgaaaaatga	5580
cccagagcgc	tgccggcacc	gttcctacga	gttgcatgat	aaagaagaca	gtcataagtg	5640
cggcgacgat	agtcatgccc	cgcgcccacc	ggaaggagct	gactgggttg	aaggctctca	5700

agggcatcgg	tcgacgctct	cccttatgcg	actcctgcat	taggaagcag	cccagtagta	5760
gggttgaggcc	ggtgagcacc	gccgcgcgcaa	ggaatgggtgc	atgcaaggag	atggcgccca	5820
acagtgcccc	ggccacgggc	ctgccaccat	acccacggcg	aaacaaggcg	tcatgagccc	5880
gaagtggcga	gcccgatctt	ccccatcggt	gatgtcggcg	atataggcg	cagcaaccgc	5940
acctgtggcg	cgggtgatgc	cggccacgat	gcgtccggcg	tagaggatct	tgccagtcac	6000
agcatgcgca	tatccatgct	tcgaccatgc	gctcacaaag	taggtgaatg	cgcaatgtag	6060
taccacacac	gtcatcgctt	tccactgctc	tcgcgaataa	agatggaaaa	tcaatctcat	6120
ggtaatagtc	catgaaaatc	cttgtattca	taaatcctcc	aggtagctat	atgcaaatgt	6180
aaacaaaaga	gatgggtgatc	tttctaagag	atgatggaat	ctcccttcag	tatcccgatg	6240
gtcaatgcgc	tggatatggg	atagatggga	atatgctgat	ttttatggga	cagagtgtcg	6300
aactgttccc	aactaaaatc	attttgcacg	atcagcgcac	tacgaacttt	accacaaaat	6360
agtcaggtaa	tgaatcctga	tataaagaca	gggttgataaa	tcagtcttct	acgcgcacatc	6420
cacgcgcaca	ccgtagaaaag	tctttcagtt	gtgagcctgg	gcaaaccggt	aactttcggc	6480
ggctttgctg	tcgcacaggc	tcacgtctaa	aaggaaataa	atcatgggtc	ataaaattat	6540
cacgttgtcc	ggcgcgccga	cggatgttct	gtatgcgctg	ttttccgctg	gcgcgttgtct	6600
gtctgggtgat	ctgccttcta	aatctggcac	agccgaattg	cgcgagcttg	gtttttgctga	6660
aaccagacac	acagaaaactg	aataccagaa	agaaaatcac	tttacctttc	tgacatcaga	6720
agggcagaaa	tttgccgttg	aacacctggg	caatacgcgt	tttgggtgagc	agcaaatattg	6780
cgcttcgatg	acgcttggcg	ttgagattga	tacctctgct	gcacaaaagg	caatcgacga	6840
gctggaccag	cgcattcgtg	acaccgtctc	cttcgaactt	attcgcaatg	gagtgctcatt	6900
catcaaggac	gccgctatcg	caaattgggtg	tatccacgca	gcggcaatcg	aaacacctca	6960
gccgggtgacc	aatatctaca	acatcagcct	tggtatccag	cgtgatgagc	cagcgcagaa	7020
caaggtaacc	gtcagtgccg	ataagttcaa	agttaaacct	gggtgtgata	ccaacattga	7080
aacgttgatc	gaaaacgcgc	tgaaaaacgc	tgctgaatgt	gcggcgctgg	atgtcacaaa	7140
gcaaatggca	gcagacaaga	aagcgatgga	tgaactggct	tcctatgtcc	gcacggccat	7200
catgatggaa	tggttccccg	gtggtgttat	ctggcagcag	tgccgtcgat	agtatgcaat	7260
tgataaattat	tatcatttgc	gggtcctttc	cggcgatccg	ccttgttacg	ggcgcgcgac	7320
ctcgcgggtt	ttcgtatatt	atgaaaattt	tccgggttaa	ggcggtttccg	ttcttcttctg	7380
tcataactta	atgtttttat	ttaaaatacc	ctctgaaaag	aaaggaaaacg	acagggtcgtt	7440
aaagcgagct	ttttggcctc	tgctgctttc	tttctctggt	tttgtccgtg	gaatgaacaa	7500
tggaagtcaa	caaaaagcag	ctgggtgaca	ttttcgggtg	gagtatccgt	accattcaga	7560
actggcagga	acagggtaatg	cccgttctgc	gagggcggtg	caagggtaat	gaggtgcttt	7620
atgactctgc	cgccgtcata	aaatgggtatg	ccgaaaggga	tgctgaaatt	gagaacgaaa	7680
agctgcgcgc	ggaggttgaa	gaactgcggc	agggccagcga	ggcagatcca	caggacgggt	7740
gtggctggca	tgactgcgta	gtcgatagtg	gctccaagta	gcgaagcgag	caggactggg	7800
cggcgccaaa	gcggctcggac	agtgtctccg	gaacgggtgc	gcatagaaat	tgcatcaacg	7860
catatagcgc	tagcagcacg	ccatagttag	tgccgatgct	gtcgggaatgg	acgatattcc	7920
gcaagaggcc	cggcagtagc	ggcataacca	agcctatgcc	tacagcatcc	aggggtgacgg	7980
tgccgaggat	gacgatgagc	gcattgttag	atttcataca	cggtgcctga	ctgcgttagc	8040
aatttaactg	tgataaacta	ccgcattaaa	gcttatcgat	gataagcggt	caaacatgag	8100
aattcgccgc	cgcaattaac	cctcactaaa	ggatcc			8136

<210> 32
 <211> 2713
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pNEB193 plasmid

<400> 32						
tcgcgcggttt	cggatgatgac	ggtgaaaacc	tctgacacat	gcagctcccg	gagacgggtca	60
cagcttgtct	gtaagcggat	gccggggagca	gacaagcccg	tcaggggcgcg	tcagcgggtg	120
ttggcggggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	180
accatattgcg	gtgtgaaata	ccgcacagat	gcgtaaggag	aaaataccgc	atcaggcgcc	240
attcgccatt	caggctgcgc	aactgttggg	aaggggcgatc	gggtgcggggc	tcttcgctat	300
tacgccagct	ggcgaaaagg	ggatgtgctg	caaggcgatt	aagttgggta	acgccagggt	360
tttcccagtc	acgacgttgt	aaaacgcacg	ccagtgaatt	cgagctcggt	acccggggggc	420
gcgcgggatc	cttaattaag	tctagagtgc	actgtttaaa	cctgcaggca	tgcaagcttg	480
gcgtaatcat	ggctcatagct	gtttcctgtg	tgaaattggt	atccgctcac	aattccacac	540
aacatacgag	ccggaagcat	aaagtgtaaa	gcctgggggtg	cctaattgagt	gagctaactc	600
acattaattg	cgtttgcgctc	actgcccgc	ttccagtcgg	gaaacctgtc	gtgccagctg	660
cattaatgaa	tcggccaacg	cgcggggaga	gagcggttgc	gtattggggc	ctcttccgct	720
tcctcgctca	ctgactcgct	gcgctcggtc	gttcggctgc	ggcgagcggt	atcagctcac	780
tcaaaggcgc	taatacgggt	atccacagaa	tcaggggata	acgcaggaaa	gaacatgtga	840
gcaaaaggcc	agcaaaaggc	caggaaaccgt	aaaaaggccg	cgttgctggc	gtttttccat	900
aggctccgcc	cccctgacga	gcatacacia	aatcgacgct	caagtcagag	gtggcgaaac	960

ccgacaggac	tataaagata	ccaggcggtt	ccccctggaa	gctccctcgt	gcgctctcct	1020
gttccgaccc	tgccgcttac	cggatacctg	tccgcctttc	tcccttcggg	aagcgtggcg	1080
ctttctcata	gctcacgctg	taggtatctc	agttcggtgt	aggtcgttcg	ctccaagctg	1140
ggctgtgtgc	acgaaccccc	cgttcagccc	gaccgctgcg	ccttatccgg	taactatcgt	1200
cttgagtcca	acccggttaag	acacgactta	tgcgccactgg	cagcagccac	tggtaacagg	1260
attagcagag	cgagggtatgt	agggcggtgct	acagagttct	tgaagtgggtg	gcctaactac	1320
ggctacacta	gaaggacagt	atcttggtatc	tgcgctctgc	tgaagccagt	taccttcgga	1380
aaaagagttg	gtagctcttg	atccggcaaaa	caaaccaccg	ctggtagcgg	tggttttttt	1440
gtttgcaagc	agcagattac	gcgcagaaaa	aaaggatctc	aagaagatcc	tttgatcttt	1500
tctacggggt	ctgacgctca	gtggaacgaa	aactcacgtt	aagggatttt	ggtcatgaga	1560
ttatcaaaaa	ggatcttcac	ctagatcctt	ttaaattaaa	aatgaagttt	taaataaatc	1620
taaagtatat	atgagtaaac	ttgggtctgac	agttaccaat	gcttaatcag	tgaggcaoct	1680
atctcagcga	tctgtctatt	tcgttcatcc	atagttgcct	gactccccgt	cgtgtagata	1740
actacgatac	gggagggcct	accatctggc	cccagtgctg	caatgatacc	gcgagaccca	1800
cgctcaccgg	ctccagattt	atcagcaata	aaccagccag	ccggaagggc	cgagcgcaga	1860
agtggctcctg	caactttatc	cgccctccatc	cagtcctatta	attggtgccg	ggaagctaga	1920
gtaagttagt	cgccagttaa	tagtttgccg	aacggttggtg	ccattgctac	aggcatcgtg	1980
gtgtcacgct	cgctcgtttg	tatggcttca	ttcagctccg	gttcccaacg	atcaaggcga	2040
gttacatgat	cccccatggt	gtgcaaaaaa	gcggttagct	ccttcggtcc	tccgatcgtt	2100
gtcagaagta	agttggccgc	agtgttatca	ctcatgggta	tggcagcact	gcataattct	2160
cttactgtca	tgccatccgt	aagatgcttt	tctgtgactg	gtgagtactc	aaccaagtca	2220
ttctgagaat	agtgtatgcg	gcgaccgagt	tgctcttgcc	cggcgtcaat	acgggataat	2280
accgcgccac	atagcagaac	tttaaaaagt	ctcatcattg	gaaaacgttc	ttcggggcga	2340
aaactctcaa	ggatcttacc	gctgttgaga	tccagttcga	tgtaaccac	tcgtgcaccc	2400
aactgatctt	cagcatcttt	tactttcacc	agcgtttctg	ggtgagcaaa	aacagggaagg	2460
caaaatgccg	caaaaaagg	aataaggggc	acacggaaat	gttgaatact	catactcttc	2520
ctttttcaat	attattgaag	catttatcag	ggttattgtc	tcatgagcgg	atacatattt	2580
gaatgtatct	agaaaaataa	acaaatagg	gttccgcgca	catttccccg	aaaagtgcc	2640
cctgacgtct	aagaaacat	tattatcatg	acattaacct	ataaaaatag	gcgtatcacg	2700
aggccctttc	gtc					2713

<210> 33
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> attP

<400> 33
 cagctttttt atactaagtt g

21

<210> 34
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> attB

<400> 34
 ctgctttttt atactaactt g

21

<210> 35
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> attL

<400> 35
 ctgctttttt atactaagtt g

21

<210> 36
 <211> 21
 <212> DNA

<213> Artificial Sequence

<220>

<223> attR

<400> 36

cagctttttt atactaactt g

21

<210> 37

<211> 1071

<212> DNA

<213> Artificial Sequence

<220>

<223> Integrase E174R

<221> CDS

<222> (1)...(1071)

<223> Nucleotide sequence encoding Integrase E147R

<400> 37

atg gga aga agg cga agt cat gag cgc cgg gat tta ccc cct aac ctt	48
Met Gly Arg Arg Arg Ser His Glu Arg Arg Asp Leu Pro Pro Asn Leu	
1 5 10 15	
tat ata aga aac aat gga tat tac tgc tac agg gac cca agg acg ggt	96
Tyr Ile Arg Asn Asn Gly Tyr Tyr Cys Tyr Arg Asp Pro Arg Thr Gly	
20 25 30	
aaa gag ttt gga tta ggc aga gac agg cga atc gca atc act gaa gct	144
Lys Glu Phe Gly Leu Gly Arg Asp Arg Arg Ile Ala Ile Thr Glu Ala	
35 40 45	
ata cag gcc aac att gag tta ttt tca gga cac aaa cac aag cct ctg	192
Ile Gln Ala Asn Ile Glu Leu Phe Ser Gly His Lys His Lys Pro Leu	
50 55 60	
aca gcg aga atc aac agt gat aat tcc gtt acg tta cat tca tgg ctt	240
Thr Ala Arg Ile Asn Ser Asp Asn Ser Val Thr Leu His Ser Trp Leu	
65 70 75 80	
gat cgc tac gaa aaa atc ctg gcc agc aga gga atc aag cag aag aca	288
Asp Arg Tyr Glu Lys Ile Leu Ala Ser Arg Gly Ile Lys Gln Lys Thr	
85 90 95	
ctc ata aat tac atg agc aaa att aaa gca ata agg agg ggt ctg cct	336
Leu Ile Asn Met Ser Lys Ile Lys Ala Ile Arg Arg Gly Leu Pro	
100 105 110	
gat gct cca ctt gaa gac atc acc aca aaa gaa att gcg gca atg ctc	384
Asp Ala Pro Leu Glu Asp Ile Thr Thr Lys Glu Ile Ala Ala Met Leu	
115 120 125	
aat gga tac ata gac gag ggc aag gcg gcg tca gcc aag tta atc aga	432
Asn Gly Tyr Ile Asp Glu Gly Lys Ala Ala Ser Ala Lys Leu Ile Arg	
130 135 140	
tca aca ctg agc gat gca ttc cga gag gca ata gct gaa ggc cat ata	480
Ser Thr Leu Ser Asp Ala Phe Arg Glu Ala Ile Ala Glu Gly His Ile	
145 150 155 160	
aca aca aac cat gtc gct gcc act cgc gca gca aaa tct aga gta agg	528
Thr Thr Asn His Val Ala Ala Thr Arg Ala Ala Lys Ser Arg Val Arg	
165 170 175	
aga tca aga ctt acg gct gac gaa tac ctg aaa att tat caa gca gca	576
Arg Ser Arg Leu Thr Ala Asp Glu Tyr Leu Lys Ile Tyr Gln Ala Ala	

180										185										190										
gaa	tca	tca	cca	tgt	tgg	ctc	aga	ctt	gca	atg	gaa	ctg	gct	gtt	gtt	624														
Glu	Ser	Ser	Pro	Cys	Trp	Leu	Arg	Leu	Ala	Met	Glu	Leu	Ala	Val	Val															
		195					200					205																		
acc	ggg	caa	cga	gtt	ggt	gat	tta	tgc	gaa	atg	aag	tgg	tct	gat	atc	672														
Thr	Gly	Gln	Arg	Val	Gly	Asp	Leu	Cys	Glu	Met	Lys	Trp	Ser	Asp	Ile															
	210					215					220																			
gta	gat	gga	tat	ctt	tat	gtc	gag	caa	agc	aaa	aca	ggc	gta	aaa	att	720														
Val	Asp	Gly	Tyr	Leu	Tyr	Val	Glu	Gln	Ser	Lys	Thr	Gly	Val	Lys	Ile															
	225				230					235					240															
gcc	atc	cca	aca	gca	ttg	cat	att	gat	gct	ctc	gga	ata	tca	atg	aag	768														
Ala	Ile	Pro	Thr	Ala	Leu	His	Ile	Asp	Ala	Leu	Gly	Ile	Ser	Met	Lys															
				245					250					255																
gaa	aca	ctt	gat	aaa	tgc	aaa	gag	att	ctt	ggc	gga	gaa	acc	ata	att	816														
Glu	Thr	Leu	Asp	Lys	Cys	Lys	Glu	Ile	Leu	Gly	Gly	Glu	Thr	Ile	Ile															
			260					265					270																	
gca	tct	act	cgt	cgc	gaa	ccg	ctt	tca	tcc	ggc	aca	gta	tca	agg	tat	864														
Ala	Ser	Thr	Arg	Arg	Glu	Pro	Leu	Ser	Ser	Gly	Thr	Val	Ser	Arg	Tyr															
		275					280					285																		
ttt	atg	cgc	gca	cga	aaa	gca	tca	ggt	ctt	tcc	ttc	gaa	ggg	gat	ccg	912														
Phe	Met	Arg	Ala	Arg	Lys	Ala	Ser	Gly	Leu	Ser	Phe	Glu	Gly	Asp	Pro															
	290					295					300																			
cct	acc	ttt	cac	gag	ttg	cgc	agt	ttg	tct	gca	aga	ctc	tat	gag	aag	960														
Pro	Thr	Phe	His	Glu	Leu	Arg	Ser	Leu	Ser	Ala	Arg	Leu	Tyr	Glu	Lys															
	305				310					315					320															
cag	ata	agc	gat	aag	ttt	gct	caa	cat	ctt	ctc	ggg	cat	aag	tcg	gac	1008														
Gln	Ile	Ser	Asp	Lys	Phe	Ala	Gln	His	Leu	Leu	Gly	His	Lys	Ser	Asp															
				325					330					335																
acc	atg	gca	tca	cag	tat	cgt	gat	gac	aga	ggc	agg	gag	tgg	gac	aaa	1056														
Thr	Met	Ala	Ser	Gln	Tyr	Arg	Asp	Asp	Arg	Gly	Arg	Glu	Trp	Asp	Lys															
		340						345					350																	
att	gaa	atc	aaa	taa												1071														
Ile	Glu	Ile	Lys	*																										
		355																												

<210> 38
 <211> 356
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Integrase E147R

<400> 38
 Met Gly Arg Arg Arg Ser His Glu Arg Arg Asp Leu Pro Pro Asn Leu
 1 5 10 15
 Tyr Ile Arg Asn Asn Gly Tyr Tyr Cys Tyr Arg Asp Pro Arg Thr Gly
 20 25 30
 Lys Glu Phe Gly Leu Gly Arg Asp Arg Arg Ile Ala Ile Thr Glu Ala
 35 40 45
 Ile Gln Ala Asn Ile Glu Leu Phe Ser Gly His Lys His Lys Pro Leu
 50 55 60
 Thr Ala Arg Ile Asn Ser Asp Asn Ser Val Thr Leu His Ser Trp Leu
 65 70 75 80
 Asp Arg Tyr Glu Lys Ile Leu Ala Ser Arg Gly Ile Lys Gln Lys Thr

[illegible]

```
<210> 39
<211> 876
<212> DNA
<213> Discosoma species
```

```
<220>
<221> CDS
<222> (45)...(737)
<223> Nucleotide sequence encoding red fluorescent
      protein (FP593)
```

<300>
<308> GenBank AF272711
<309> 2000-09-26

```
<400> 39
agtttcagcc agtgacaggg tgagctgcc a ggtattctaa caag atg agt tgt tcc 56
                                     Met Ser Cys Ser
                                     1
```

aag aat gtg atc aag gag ttc atg agg ttc aag gtt cgt atg gaa gga 104
lys asn val ile lys glu phe met arg phe lys val arg met glu gly
5 10 15 20

acg gtc aat ggg cac gag ttt gaa ata aaa ggc gaa ggt gaa ggg agg 152
Thr Val Asn Gly His Glu Phe Glu Ile Lys Gly Glu Gly Glu Gly Arg
25 30 35

cct tac gaa ggt cac tgt tcc gta aag ctt atg gta acc aag ggt gga 200
Pro Tyr Glu Gly His Cys Ser Val Lys Leu Met Val Thr Lys Gly Gly

40										45										50										
cct	ttg	cca	ttt	gct	ttt	gat	att	ttg	tca	cca	caa	ttt	cag	tat	gga	248														
Pro	Leu	Pro	Phe	Ala	Phe	Asp	Ile	Leu	Ser	Pro	Gln	Phe	Gln	Tyr	Gly															
		55					60					65																		
agc	aag	gta	tat	gtc	aaa	cac	cct	gcc	gac	ata	cca	gac	tat	aaa	aag	296														
Ser	Lys	Val	Tyr	Val	Lys	His	Pro	Ala	Asp	Ile	Pro	Asp	Tyr	Lys	Lys															
	70					75					80																			
ctg	tca	ttt	cct	gag	gga	ttt	aaa	tgg	gaa	agg	gtc	atg	aac	ttt	gaa	344														
Leu	Ser	Phe	Pro	Glu	Gly	Phe	Lys	Trp	Glu	Arg	Val	Met	Asn	Phe	Glu															
	85				90					95					100															
gac	ggt	ggc	gtg	gtt	act	gta	tcc	caa	gat	tcc	agt	ttg	aaa	gac	ggc	392														
Asp	Gly	Gly	Val	Val	Thr	Val	Ser	Gln	Asp	Ser	Ser	Leu	Lys	Asp	Gly															
				105					110					115																
tgt	ttc	atc	tac	gag	gtc	aag	ttc	att	ggg	gtg	aac	ttt	cct	tct	gat	440														
Cys	Phe	Ile	Tyr	Glu	Val	Lys	Phe	Ile	Gly	Val	Asn	Phe	Pro	Ser	Asp															
			120					125					130																	
gga	cct	gtt	atg	cag	agg	agg	aca	cgg	ggc	tgg	gaa	gcc	agc	tct	gag	488														
Gly	Pro	Val	Met	Gln	Arg	Arg	Thr	Arg	Gly	Trp	Glu	Ala	Ser	Ser	Glu															
		135					140					145																		
cgt	ttg	tat	cct	cgt	gat	ggg	gtg	ctg	aaa	gga	gac	atc	cat	atg	gct	536														
Arg	Leu	Tyr	Pro	Arg	Asp	Gly	Val	Leu	Lys	Gly	Asp	Ile	His	Met	Ala															
	150				155						160																			
ctg	agg	ctg	gaa	gga	ggc	ggc	cat	tac	ctc	gtt	gaa	ttc	aaa	agt	att	584														
Leu	Arg	Leu	Glu	Gly	Gly	Gly	His	Tyr	Leu	Val	Glu	Phe	Lys	Ser	Ile															
	165				170					175					180															
tac	atg	gta	aag	aag	cct	tca	gtg	cag	ttg	cca	ggc	tac	tat	tat	gtt	632														
Tyr	Met	Val	Lys	Lys	Pro	Ser	Val	Gln	Leu	Pro	Gly	Tyr	Tyr	Tyr	Val															
				185					190					195																
gac	tcc	aaa	ctg	gat	atg	acg	agc	cac	aac	gaa	gat	tac	aca	gtc	gtt	680														
Asp	Ser	Lys	Leu	Asp	Met	Thr	Ser	His	Asn	Glu	Asp	Tyr	Thr	Val	Val															
			200					205					210																	
gag	cag	tat	gaa	aaa	acc	cag	gga	cgc	cac	cat	ccg	ttc	att	aag	cct	728														
Glu	Gln	Tyr	Glu	Lys	Thr	Gln	Gly	Arg	His	His	Pro	Phe	Ile	Lys	Pro															
		215					220					225																		
ctg	cag	tga	actcggtctca	gtcatggatt	agcggtaatg	gccacaaaag										777														
Leu	Gln	*																												
	230																													
gcacgatgat	cgttttttag	gaatgcagcc	aaaaattgaa	ggttatgaca	gtagaaatac											837														
aagcaacagg	ctttgcttat	taaacatgta	attgaaaac													876														

<210> 40
 <211> 230
 <212> PRT
 <213> Discosoma species

<400> 40
 Met Ser Cys Ser Lys Asn Val Ile Lys Glu Phe Met Arg Phe Lys Val
 1 5 10 15
 Arg Met Glu Gly Thr Val Asn Gly His Glu Phe Glu Ile Lys Gly Glu
 20 25 30
 Gly Glu Gly Arg Pro Tyr Glu Gly His Cys Ser Val Lys Leu Met Val
 35 40 45
 Thr Lys Gly Gly Pro Leu Pro Phe Ala Phe Asp Ile Leu Ser Pro Gln
 50 55 60

-25-

Phe	Gln	Tyr	Gly	Ser	Lys	Val	Tyr	Val	Lys	His	Pro	Ala	Asp	Ile	Pro
65					70					75					80
Asp	Tyr	Lys	Lys	Leu	Ser	Phe	Pro	Glu	Gly	Phe	Lys	Trp	Glu	Arg	Val
				85					90					95	
Met	Asn	Phe	Glu	Asp	Gly	Gly	Val	Val	Thr	Val	Ser	Gln	Asp	Ser	Ser
			100					105					110		
Leu	Lys	Asp	Gly	Cys	Phe	Ile	Tyr	Glu	Val	Lys	Phe	Ile	Gly	Val	Asn
			115				120					125			
Phe	Pro	Ser	Asp	Gly	Pro	Val	Met	Gln	Arg	Arg	Thr	Arg	Gly	Trp	Glu
	130					135					140				
Ala	Ser	Ser	Glu	Arg	Leu	Tyr	Pro	Arg	Asp	Gly	Val	Leu	Lys	Gly	Asp
145					150					155					160
Ile	His	Met	Ala	Leu	Arg	Leu	Glu	Gly	Gly	His	Tyr	Leu	Val	Glu	
				165					170					175	
Phe	Lys	Ser	Ile	Tyr	Met	Val	Lys	Lys	Pro	Ser	Val	Gln	Leu	Pro	Gly
			180					185					190		
Tyr	Tyr	Tyr	Val	Asp	Ser	Lys	Leu	Asp	Met	Thr	Ser	His	Asn	Glu	Asp
		195					200					205			
Tyr	Thr	Val	Val	Glu	Gln	Tyr	Glu	Lys	Thr	Gln	Gly	Arg	His	His	Pro
	210					215					220				
Phe	Ile	Lys	Pro	Leu	Gln										
225					230										

<210> 41
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> m-att;

<221> misc_difference
 <222> 18
 <223> n is a or g or c or t/u

<400> 41
 rkycwgcttt yktrtacnaa stsgb

25

<210> 42
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> m-attB;

<221> misc_difference
 <222> 18
 <223> n is a or g or c or t/u

<400> 42
 agccwgcttt yktrtacnaa ctsgb

25

<210> 43
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> m-attR

<221> misc_difference
 <222> 18
 <223> n is a or g or c or t/u

<400> 43
gttcagcttt cktrtacnaa ctsgb 25

<210> 44
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> m-attI

<221> misc_difference
<222> 18
<223> n is a or g or c or t/u

<400> 44
agccwgcttt cktrtacnaa gtsgb 25

<210> 45
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> m-attP1

<221> misc_difference
<222> 18
<223> n is a or g or c or t/u

<400> 45
gttcagcttt yktrtacnaa gtsgb 25

<210> 46
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attB1

<400> 46
agcctgcttt tttgtacaaa cttgt 25

<210> 47
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attB2

<400> 47
agcctgcttt cttgtacaaa cttgt 25

<210> 48
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attB3

<400> 48
accagcttt cttgtacaaa cttgt 25

<210> 49

<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attR1

<400> 49
gttcagcttt tttgtacaaa cttgt 25

<210> 50
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attR2

<400> 50
gttcagcttt cttgtacaaa cttgt 25

<210> 51
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attR3

<400> 51
gttcagcttt cttgtacaaa gttgg 25

<210> 52
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attL1

<400> 52
agcctgcttt tttgtacaaa gttgg 25

<210> 53
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attL2

<400> 53
agcctgcttt cttgtacaaa gttgg 25

<210> 54
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attL3

<400> 54
acccagcttt cttgtacaaa gttgg 25

<210> 55
<211> 25


```

<212> DNA
<213> Artificial Sequence

<220>
<223> attP1

<400> 55
gttcagcttt tttgtacaaa gttgg                25

<210> 56
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attP2,P3

<400> 56
gttcagcttt cttgtacaaa gttgg                25

<210> 57
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Lox P site

<400> 57
ataacttcgt ataatgtatg ctatacgaag ttat        34

<210> 58
<211> 1032
<212> DNA
<213> Escherichia coli

<220>
<221> CDS
<222> (1)...(1032)
<223> nucleotide sequence encoding Cre recombinase

<400> 58
atg tcc aat tta ctg acc gta cac caa aat ttg cct gca tta ccg gtc        48
Met Ser Asn Leu Leu Thr Val His Gln Asn Leu Pro Ala Leu Pro Val
  1             5             10             15

gat gca acg agt gat gag gtt cgc aag aac ctg atg gac atg ttc agg        96
Asp Ala Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Met Phe Arg
          20             25             30

gat cgc cag gcg ttt tct gag cat acc tgg aaa atg ctt ctg tcc gtt        144
Asp Arg Gln Ala Phe Ser Glu His Thr Trp Lys Met Leu Leu Ser Val
          35             40             45

tgc cgg tcg tgg gcg gca tgg tgc aag ttg aat aac cgg aaa tgg ttt        192
Cys Arg Ser Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Trp Phe
          50             55             60

ccc gca gaa cct gaa gat gtt cgc gat tat ctt cta tat ctt cag gcg        240
Pro Ala Glu Pro Glu Val Arg Asp Tyr Leu Leu Tyr Leu Gln Ala
          65             70             75             80

cgc ggt ctg gca gta aaa act atc cag caa cat ttg ggc cag cta aac        288
Arg Gly Leu Ala Val Lys Thr Ile Gln Gln His Leu Gly Gln Leu Asn
          85             90             95

atg ctt cat cgt cgg tcc ggg ctg cca cga cca agt gac agc aat gct        336

```

Met	Leu	His	Arg	Arg	Ser	Gly	Leu	Pro	Arg	Pro	Ser	Asp	Ser	Asn	Ala		
			100					105					110				
gtt	tca	ctg	gtt	atg	cgg	cgg	atc	cga	aaa	gaa	aac	gtt	gat	gcc	ggg	384	
Val	Ser	Leu	Val	Met	Arg	Arg	Ile	Arg	Lys	Glu	Asn	Val	Asp	Ala	Gly		
		115					120					125					
gaa	cgt	gca	aaa	cag	gct	cta	gcg	ttc	gaa	cgc	act	gat	ttc	gac	cag	432	
Glu	Arg	Ala	Lys	Gln	Ala	Leu	Ala	Phe	Glu	Arg	Thr	Asp	Phe	Asp	Gln		
	130					135					140						
gtt	cgt	tca	ctc	atg	gaa	aat	agc	gat	cgc	tgc	cag	gat	ata	cgt	aat	480	
Val	Arg	Ser	Leu	Met	Glu	Asn	Ser	Asp	Arg	Cys	Gln	Asp	Ile	Arg	Asn		
	145				150					155					160		
ctg	gca	ttt	ctg	ggg	att	gct	tat	aac	acc	ctg	tta	cgt	ata	gcc	gaa	528	
Leu	Ala	Phe	Leu	Gly	Ile	Ala	Tyr	Asn	Thr	Leu	Leu	Arg	Ile	Ala	Glu		
				165					170					175			
att	gcc	agg	atc	agg	gtt	aaa	gat	atc	tca	cgt	act	gac	ggg	ggg	aga	576	
Ile	Ala	Arg	Ile	Arg	Val	Lys	Asp	Ile	Ser	Arg	Thr	Asp	Gly	Gly	Arg		
		180					185						190				
atg	tta	atc	cat	att	ggc	aga	acg	aaa	acg	ctg	gtt	agc	acc	gca	ggg	624	
Met	Leu	Ile	His	Ile	Gly	Arg	Thr	Lys	Thr	Leu	Val	Ser	Thr	Ala	Gly		
		195				200						205					
gta	gag	aag	gca	ctt	agc	ctg	ggg	gta	act	aaa	ctg	gtc	gag	cga	tgg	672	
Val	Glu	Lys	Ala	Leu	Ser	Gly	Gly	Val	Thr	Lys	Leu	Val	Glu	Arg	Trp		
	210					215					220						
att	tcc	gtc	tct	ggg	gta	gct	gat	gat	ccg	aat	aac	tac	ctg	ttt	tgc	720	
Ile	Ser	Val	Ser	Gly	Val	Ala	Asp	Asp	Pro	Asn	Asn	Tyr	Leu	Phe	Cys		
	225				230				235						240		
cgg	gtc	aga	aaa	aat	ggg	gtt	gcc	gcg	cca	tct	gcc	acc	agc	cag	cta	768	
Arg	Val	Arg	Lys	Asn	Gly	Val	Ala	Ala	Pro	Ser	Ala	Thr	Ser	Gln	Leu		
				245					250					255			
tca	act	cgc	gcc	ctg	gaa	ggg	att	ttt	gaa	gca	act	cat	cga	ttg	att	816	
Ser	Thr	Arg	Ala	Leu	Glu	Gly	Ile	Phe	Glu	Ala	Thr	His	Arg	Leu	Ile		
			260					265					270				
tac	ggc	gct	aag	gat	gac	tct	ggg	cag	aga	tac	ctg	gcc	tgg	tct	gga	864	
Tyr	Gly	Ala	Lys	Asp	Asp	Ser	Gly	Gln	Arg	Tyr	Leu	Ala	Trp	Ser	Gly		
		275					280					285					
cac	agt	gcc	cgt	gtc	gga	gcc	gcg	cga	gat	atg	gcc	cgc	gct	gga	gtt	912	
His	Ser	Ala	Arg	Val	Gly	Ala	Ala	Arg	Asp	Met	Ala	Arg	Ala	Gly	Val		
	290				295						300						
tca	ata	ccg	gag	atc	atg	caa	gct	ggg	ggc	tgg	acc	aat	gta	aat	att	960	
Ser	Ile	Pro	Glu	Ile	Met	Gln	Ala	Gly	Gly	Trp	Thr	Asn	Val	Asn	Ile		
	305				310					315					320		
gtc	atg	aac	tat	atc	cgt	aac	ctg	gat	agt	gaa	aca	ggg	gca	atg	gtg	1008	
Val	Met	Asn	Tyr	Ile	Arg	Asn	Leu	Asp	Ser	Glu	Thr	Gly	Ala	Met	Val		
				325					330					335			
cgc	ctg	ctg	gaa	gat	ggc	gat	tag									1032	
Arg	Leu	Leu	Glu	Asp	Gly	Asp	*										
			340														

<210> 59
 <211> 343
 <212> PRT

<213> Escherichia coli

<400> 59

```

Met Ser Asn Leu Leu Thr Val His Gln Asn Leu Pro Ala Leu Pro Val
1      5      10      15
Asp Ala Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Met Phe Arg
      20      25      30
Asp Arg Gln Ala Phe Ser Glu His Thr Trp Lys Met Leu Leu Ser Val
      35      40      45
Cys Arg Ser Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Trp Phe
      50      55      60
Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu Tyr Leu Gln Ala
      65      70      75      80
Arg Gly Leu Ala Val Lys Thr Ile Gln Gln His Leu Gly Gln Leu Asn
      85      90      95
Met Leu His Arg Arg Ser Gly Leu Pro Arg Pro Ser Asp Ser Asn Ala
      100      105      110
Val Ser Leu Val Met Arg Arg Ile Arg Lys Glu Asn Val Asp Ala Gly
      115      120      125
Glu Arg Ala Lys Gln Ala Leu Ala Phe Glu Arg Thr Asp Phe Asp Gln
      130      135      140
Val Arg Ser Leu Met Glu Asn Ser Asp Arg Cys Gln Asp Ile Arg Asn
      145      150      155      160
Leu Ala Phe Leu Gly Ile Ala Tyr Asn Thr Leu Leu Arg Ile Ala Glu
      165      170      175
Ile Ala Arg Ile Arg Val Lys Asp Ile Ser Arg Thr Asp Gly Gly Arg
      180      185      190
Met Leu Ile His Ile Gly Arg Thr Lys Thr Leu Val Ser Thr Ala Gly
      195      200      205
Val Glu Lys Ala Leu Ser Leu Gly Val Thr Lys Leu Val Glu Arg Trp
      210      215      220
Ile Ser Val Ser Gly Val Ala Asp Asp Pro Asn Asn Tyr Leu Phe Cys
      225      230      235      240
Arg Val Arg Lys Asn Gly Val Ala Ala Pro Ser Ala Thr Ser Gln Leu
      245      250      255
Ser Thr Arg Ala Leu Glu Gly Ile Phe Glu Ala Thr His Arg Leu Ile
      260      265      270      275
Tyr Gly Ala Lys Asp Asp Ser Gly Gln Arg Tyr Leu Ala Trp Ser Gly
      275      280      285
His Ser Ala Arg Val Gly Ala Ala Arg Asp Met Ala Arg Ala Gly Val
      290      295      300
Ser Ile Pro Glu Ile Met Gln Ala Gly Gly Trp Thr Asn Val Asn Ile
      305      310      315      320
Val Met Asn Tyr Ile Arg Asn Leu Asp Ser Glu Thr Gly Ala Met Val
      325      330      335
Arg Leu Leu Glu Asp Gly Asp
      340

```

<210> 60

<211> 1272

<212> DNA

<213> Saccharomyces cerevisiae

<220>

<221> CDS

<222> (1)...(1272)

<223> nucleotide sequence encoding Flip recombinase

<400> 60

```

atg cca caa ttt ggt ata tta tgt aaa aca cca cct aag gtg ctt gtt
Met Pro Gln Phe Gly Ile Leu Cys Lys Thr Pro Pro Lys Val Leu Val
1      5      10      15

cgt cag ttt gtg gaa agg ttt gaa aga cct tca ggt gag aaa ata gca
Arg Gln Phe Val Glu Arg Phe Glu Arg Pro Ser Gly Glu Lys Ile Ala
      20      25      30

```

48

96

-31-

tta tgt gct gct gaa cta acc tat tta tgt tgg atg att aca cat aac	144
Leu Cys Ala Ala Glu Leu Thr Tyr Leu Cys Trp Met Ile Thr His Asn	
35 40 45	
gga aca gca atc aag aga gcc aca ttc atg agc tat aat act atc ata	192
Gly Thr Ala Ile Lys Arg Ala Thr Phe Met Ser Tyr Asn Thr Ile Ile	
50 55 60	
agc aat tgc ctg agt ttc gat att gtc aat aaa tca ctc cag ttt aaa	240
Ser Asn Ser Leu Ser Phe Asp Ile Val Asn Lys Ser Leu Gln Phe Lys	
65 70 75 80	
tac aag acg caa aaa gca aca att ctg gaa gcc tca tta aag aaa ttg	288
Tyr Lys Thr Gln Lys Ala Thr Ile Leu Glu Ala Ser Leu Lys Lys Leu	
85 90 95	
att cct gct tgg gaa ttt aca att att cct tac tat gga caa aaa cat	336
Ile Pro Ala Trp Glu Phe Thr Ile Ile Pro Tyr Tyr Gly Gln Lys His	
100 105 110	
caa tct gat atc act gat att gta agt agt ttg caa tta cag ttc gaa	384
Gln Ser Asp Ile Thr Asp Ile Val Ser Ser Leu Gln Leu Gln Phe Glu	
115 120 125	
tca tgc gaa gaa gca gat aag gga aat agc cac agt aaa aaa atg ctt	432
Ser Ser Glu Glu Ala Asp Lys Gly Asn Ser His Ser Lys Lys Met Leu	
130 135 140	
aaa gca ctt cta agt gag ggt gaa agc atc tgg gag atc act gag aaa	480
Lys Ala Leu Leu Ser Glu Gly Glu Ser Ile Trp Glu Ile Thr Glu Lys	
145 150 155 160	
ata cta aat tgc ttt gag tat act tgc aga ttt aca aaa aca aaa act	528
Ile Leu Asn Ser Phe Glu Tyr Thr Ser Arg Phe Thr Lys Thr Lys Thr	
165 170 175	
tta tac caa ttc ctc ttc cta gct act ttc atc aat tgt gga aga ttc	576
Leu Tyr Gln Phe Leu Phe Leu Ala Thr Phe Ile Asn Cys Gly Arg Phe	
180 185 190	
agc gat att aag aac gtt gat ccg aaa tca ttt aaa tta gtc caa aat	624
Ser Asp Ile Lys Asn Val Asp Pro Lys Ser Phe Lys Leu Val Gln Asn	
195 200 205	
aag tat ctg gga gta ata atc cag tgt tta gtg aca gag aca aag aca	672
Lys Tyr Leu Gly Val Ile Ile Gln Cys Leu Val Thr Glu Thr Lys Thr	
210 215 220	
agc gtt agt agg cac ata tac ttc ttt agc gca agg ggt agg atc gat	720
Ser Val Ser Arg His Ile Tyr Phe Phe Ser Ala Arg Gly Arg Ile Asp	
225 230 235 240	
cca ctt gta tat ttg gat gaa ttt ttg agg aat tct gaa cca gtc cta	768
Pro Leu Val Tyr Leu Asp Glu Phe Leu Arg Asn Ser Glu Pro Val Leu	
245 250 255	
aaa cga gta aat agg acc ggc aat tct tca agc aat aaa cag gaa tac	816
Lys Arg Val Asn Arg Thr Gly Asn Ser Ser Ser Asn Lys Gln Glu Tyr	
260 265 270	
caa tta tta aaa gat aac tta gtc aga tgc tac aat aaa gct ttg aag	864
Gln Leu Leu Lys Asp Asn Leu Val Arg Ser Tyr Asn Lys Ala Leu Lys	
275 280 285	
aaa aat gcg cct tat tca atc ttt gct ata aaa aat ggc cca aaa tct	912
Lys Asn Ala Pro Tyr Ser Ile Phe Ala Ile Lys Asn Gly Pro Lys Ser	
290 295 300	

cac	att	gga	aga	cat	ttg	atg	acc	tca	ttt	ctt	tca	atg	aag	ggc	cta	960
His	Ile	Gly	Arg	His	Leu	Met	Thr	Ser	Phe	Leu	Ser	Met	Lys	Gly	Leu	
305					310					315					320	
acg	gag	ttg	act	aat	gtt	gtg	gga	aat	tgg	agc	gat	aag	cgt	gct	tct	1008
Thr	Glu	Leu	Thr	Asn	Val	Val	Gly	Asn	Trp	Ser	Asp	Lys	Arg	Ala	Ser	
				325					330					335		
gcc	gtg	gcc	agg	aca	acg	tat	act	cat	cag	ata	aca	gca	ata	cct	gat	1056
Ala	Val	Ala	Arg	Thr	Thr	Tyr	Thr	His	Gln	Ile	Thr	Ala	Ile	Pro	Asp	
			340					345					350			
cac	tac	ttc	gca	cta	gtt	tct	cgg	tac	tat	gca	tat	gat	cca	ata	tca	1104
His	Tyr	Phe	Ala	Leu	Val	Ser	Arg	Tyr	Tyr	Ala	Tyr	Asp	Pro	Ile	Ser	
		355					360					365				
aag	gaa	atg	ata	gca	ttg	aag	gat	gag	act	aat	cca	att	gag	gag	tgg	1152
Lys	Glu	Met	Ile	Ala	Leu	Lys	Asp	Glu	Thr	Asn	Pro	Ile	Glu	Glu	Trp	
	370					375					380					
cag	cat	ata	gaa	cag	cta	aag	ggt	agt	gct	gaa	gga	agc	ata	cga	tac	1200
Gln	His	Ile	Glu	Gln	Leu	Lys	Gly	Ser	Ala	Glu	Gly	Ser	Ile	Arg	Tyr	
385					390					395					400	
ccc	gca	tgg	aat	ggg	ata	ata	tca	cag	gag	gta	cta	gac	tac	ctt	tca	1248
Pro	Ala	Trp	Asn	Gly	Ile	Ile	Ser	Gln	Glu	Val	Leu	Asp	Tyr	Leu	Ser	
				405					410					415		
tcc	tac	ata	aat	aga	cgc	ata	taa									1272
Ser	Tyr	Ile	Asn	Arg	Arg	Ile	*									
			420													

<210> 61
 <211> 422
 <212> PRT
 <213> Saccharomyces cerevisiae

<400> 61
 Pro Gln Phe Gly Ile Leu Cys Lys Thr Pro Pro Lys Val Leu Val Arg
 1 5 10 15
 Gln Phe Val Glu Arg Phe Glu Arg Pro Ser Gly Glu Lys Ile Ala Leu
 20 25 30
 Cys Ala Ala Glu Leu Thr Tyr Leu Cys Trp Met Ile Thr His Asn Gly
 35 40 45
 Thr Ala Ile Lys Arg Ala Thr Phe Met Ser Tyr Asn Thr Ile Ile Ser
 50 55 60
 Asn Ser Leu Ser Phe Asp Ile Val Asn Lys Ser Leu Gln Phe Lys Tyr
 65 70 75 80
 Lys Thr Gln Lys Ala Thr Ile Leu Glu Ala Ser Leu Lys Lys Leu Ile
 85 90 95
 Pro Ala Trp Glu Phe Thr Ile Ile Pro Tyr Tyr Gly Gln Lys His Gln
 100 105 110
 Ser Asp Ile Thr Asp Ile Val Ser Ser Leu Gln Leu Gln Phe Glu Ser
 115 120 125
 Ser Glu Glu Ala Asp Lys Gly Asn Ser His Ser Lys Lys Met Leu Lys
 130 135 140
 Ala Leu Leu Ser Glu Gly Glu Ser Ile Trp Glu Ile Thr Glu Lys Ile
 145 150 155 160
 Leu Asn Ser Phe Glu Tyr Thr Ser Arg Phe Thr Lys Thr Lys Thr Leu
 165 170 175
 Tyr Gln Phe Leu Phe Leu Ala Thr Phe Ile Asn Cys Gly Arg Phe Ser
 180 185 190
 Asp Ile Lys Asn Val Asp Pro Lys Ser Phe Lys Leu Val Gln Asn Lys
 195 200 205
 Tyr Leu Gly Val Ile Ile Gln Cys Leu Val Thr Glu Thr Lys Thr Ser

210	Val	Ser	Arg	His	Ile	Tyr	Phe	Phe	Ser	Ala	Arg	Gly	Arg	Ile	Asp	Pro
225	Leu	Val	Tyr	Leu	Asp	Glu	Phe	Leu	Arg	Asn	Ser	Glu	Pro	Val	Leu	Lys
	Arg	Val	Asn	Arg	Thr	Gly	Asn	Ser	Ser	Ser	Asn	Lys	Gln	Glu	Tyr	Gln
	Leu	Leu	Lys	Asp	Asn	Leu	Val	Arg	Ser	Tyr	Asn	Lys	Ala	Leu	Lys	Lys
	Asn	Ala	Pro	Tyr	Ser	Ile	Phe	Ala	Ile	Lys	Asn	Gly	Pro	Lys	Ser	His
	Ile	Gly	Arg	His	Leu	Met	Thr	Ser	Phe	Leu	Ser	Met	Lys	Gly	Leu	Thr
	Glu	Leu	Thr	Asn	Val	Val	Gly	Asn	Trp	Ser	Asp	Lys	Arg	Ala	Ser	Ala
	Val	Ala	Arg	Thr	Thr	Tyr	Thr	His	Gln	Ile	Thr	Ala	Ile	Pro	Asp	His
	Tyr	Phe	Ala	Leu	Val	Ser	Arg	Tyr	Tyr	Ala	Tyr	Asp	Pro	Ile	Ser	Lys
	Glu	Met	Ile	Ala	Leu	Lys	Asp	Glu	Thr	Asn	Pro	Ile	Glu	Glu	Trp	Gln
	His	Ile	Glu	Gln	Leu	Lys	Gly	Ser	Ala	Glu	Gly	Ser	Ile	Arg	Tyr	Pro
	Ala	Trp	Asn	Gly	Ile	Ile	Ser	Gln	Glu	Val	Leu	Asp	Tyr	Leu	Ser	Ser
	Tyr	Ile	Asn	Arg	Arg	Ile										

<210> 62
 <211> 48
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> IR2

<400> 62
 gaagttccta ttccgaagtt cctattctct agaaagtata ggaacttc

48

<210> 63
 <211> 48
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> IR1

<400> 63
 gaagttccta tacttttctag agaataggaa cttcgggaata ggaacttc

48

<210> 64
 <211> 66
 <212> DNA
 <213> Bacteriophage mu

<220>
 <221> CDS
 <222> (1)...(66)
 <223> nucleotide sequence encoding GIN recombinase

<400> 64	tca	act	ctg	tat	aaa	aaa	cac	ccc	gcg	aaa	cga	gcg	cat	ata	gaa	aac
	Ser	Thr	Leu	Tyr	Lys	Lys	His	Pro	Ala	Lys	Arg	Ala	His	Ile	Glu	Asn
	1				5				10					15		

48

gac gat cga atc aat taa
 Asp Asp Arg Ile Asn *

66

20

<210> 65
 <211> 21
 <212> PRT
 <213> bacteriophage mu

<400> 65
 Ser Thr Leu Tyr Lys Lys His Pro Ala Lys Arg Ala His Ile Glu Asn
 1 5 10 15
 Asp Asp Arg Ile Asn
 20

<210> 66
 <211> 69
 <212> DNA
 <213> Bacteriophage mu

<220>
 <221> CDS
 <222> (1)...(69)
 <223> nucleotide sequence encoding Gin recombinase

<400> 66
 tat aaa aaa cat ccc gcg aaa cga acg cat ata gaa aac gac gat cga 48
 Tyr Lys Lys His Pro Ala Lys Arg Thr His Ile Glu Asn Asp Asp Arg
 1 5 10 15

atc aat caa atc gat cgg taa 69
 Ile Asn Gln Ile Asp Arg *
 20

<210> 67
 <211> 22
 <212> PRT
 <213> bacteriophage mu

<220>
 <223> Gin recombinase of bacteriophage mu

<400> 67
 Tyr Lys Lys His Pro Ala Lys Arg Thr His Ile Glu Asn Asp Asp Arg
 1 5 10 15
 Ile Asn Gln Ile Asp Arg
 20

<210> 68
 <211> 555
 <212> DNA
 <213> Escherichia coli

<220>
 <221> CDS
 <222> (1)...(555)
 <223> nucleotide sequence encoding PIN recombinase

<400> 68
 atg ctt att ggc tat gta cgc gta tca aca aat gac cag aac aca gat 48
 Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
 1 5 10 15

cta caa cgt aat gcg ctg aac tgt gca gga tgc gag ctg att ttt gaa 96
 Leu Gln Arg Asn Ala Leu Asn Cys Ala Gly Cys Glu Leu Ile Phe Glu
 20 25 30

-35-

gac aag ata agc ggc aca aag tcc gaa agg ccg gga ctg aaa aaa ctg	144
Asp Lys Ile Ser Gly Thr Lys Ser Glu Arg Pro Gly Leu Lys Lys Leu	
35 40 45	
ctc agg aca tta tgc gca ggt gac act ctg gtt gtc tgg aag ctg gat	192
Leu Arg Thr Leu Ser Ala Gly Asp Thr Leu Val Val Trp Lys Leu Asp	
50 55 60	
cgg ctg ggg cgt agt atg cgg cat ctt gtc gtg ctg gtg gag gag ttg	240
Arg Leu Gly Arg Ser Met Arg His Leu Val Val Leu Val Glu Glu Leu	
65 70 75 80	
cgc gaa cga ggc atc aac ttt cgt agt ctg acg gat tca att gat acc	288
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr	
85 90 95	
agc aca cca atg gga cgc ttt ttc ttt cat gtg atg ggt gcc ctg gct	336
Ser Thr Pro Met Gly Arg Phe Phe Phe His Val Met Gly Ala Leu Ala	
100 105 110	
gaa atg gag cgt gaa ctg att gtt gaa cga aca aaa gct gga ctg gaa	384
Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Lys Ala Gly Leu Glu	
115 120 125	
act gct cgt gca cag gga cga att ggt gga cgt cgt ccc aaa ctt aca	432
Thr Ala Arg Ala Gln Gly Arg Ile Gly Gly Arg Arg Pro Lys Leu Thr	
130 135 140	
cca gaa caa tgg gca caa gct gga cga tta att gca gca gga act cct	480
Pro Glu Gln Trp Ala Gln Ala Gly Arg Leu Ile Ala Ala Gly Thr Pro	
145 150 155 160	
cgc cag aag gtg gcg att atc tat gat gtt ggt gtg tca act ttg tat	528
Arg Gln Lys Val Ala Ile Ile Tyr Asp Val Gly Val Ser Thr Leu Tyr	
165 170 175	
aag agg ttt cct gca ggg gat aaa taa	555
Lys Arg Phe Pro Ala Gly Asp Lys *	
180	

<210> 69
 <211> 184
 <212> PRT
 <213> Escherichia coli

<400> 69
Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15
Leu Gln Arg Asn Ala Leu Asn Cys Ala Gly Cys Glu Leu Ile Phe Glu
20 25 30
Asp Lys Ile Ser Gly Thr Lys Ser Glu Arg Pro Gly Leu Lys Lys Leu
35 40 45
Leu Arg Thr Leu Ser Ala Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60
Arg Leu Gly Arg Ser Met Arg His Leu Val Val Leu Val Glu Glu Leu
65 70 75 80
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95
Ser Thr Pro Met Gly Arg Phe Phe Phe His Val Met Gly Ala Leu Ala
100 105 110
Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Lys Ala Gly Leu Glu
115 120 125
Thr Ala Arg Ala Gln Gly Arg Ile Gly Gly Arg Arg Pro Lys Leu Thr
130 135 140
Pro Glu Gln Trp Ala Gln Ala Gly Arg Leu Ile Ala Ala Gly Thr Pro
145 150 155 160

Arg Gln Lys Val Ala Ile Ile Tyr Asp Val Gly Val Ser Thr Leu Tyr
 165 170 175
 Lys Arg Phe Pro Ala Gly Asp Lys
 180

<210> 70
 <211> 4778
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pcx plasmid

<400> 70
 gtcgacattg attattgact agttattaat agtaatcaat tacgggggtca ttagttcata 60
 gcccataatat ggagttccgc gttacataac ttacgggtaaa tggcccgccct ggctgaccgc 120
 ccaacgaccc cgcgccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180
 ggactttcca ttgacgtcaa tgggtggact atttacggta aactgcccac ttggcagtac 240
 atcaagtgtg tcatatgcca agtacgcccc ctattgacgt caatgacggg aaatggcccg 300
 cctggcatta tgcccagtac atgaccttat gggactttcc tacttggcag tacatctacg 360
 tattagtcac cgctattacc atgggtcgag gtgagcccca cgttctgctt cactctcccc 420
 atctccccc cctccccacc cccaattttg tatttattta ttttttaatt attttgtgca 480
 gcgatggggg cggggggggg ggggggcgcg gccaggcggg gcggggcggg gcgagggggc 540
 gggcgggggc aggcggagag gtgcggcgcc agccaatcag agcggcgccg tccgaaagt 600
 tccttttatg gcgagggcgg ggcggcgggc gccctataaa aagcgaagcg cgcggcgggc 660
 gggagtcgct gcgttgccct cgcgccgtgc cccgctccgc gccgctcgc gccgcccgcc 720
 ccgctctga ctgaccgctg tactccaca ggtgagcggg cgggacggcc cttctctcc 780
 gggctgtaac tagcgcttgg ttaatgacg gctcgcttct tttctgtggc tgcgtgaaag 840
 ccttaaaggc ctccgggagg gccctttgtg cggggggggg cggctcgggg ggtgcgtgcg 900
 tgtgtgtgtg cgtggggagg gcccgctgcg gcccgcggtg cccggcggtg gtgagcgctg 960
 cgggcgcggc gcggggcctt gtgcgctccg cgtgtgcgcg aggggagcgc ggccgggggc 1020
 ggtgcccgcg ggtgcggggg ggctgcgagg ggaacaaagg ctgcgtgcgg ggtgtgtgcg 1080
 tggggggagg agcagggggg gtgcggcgcg cggctcgggc gtaaccccc cctgcacccc 1140
 cctccccgag ttgctgagca cggcccggtc tcgggtgcgg ggctccgtgc ggggcgtggc 1200
 gcggggctcg ccgtgcccgg cgggggggtg cggcaggtgg ggggtgccgg cggggcgggg 1260
 ccgctcggg cgggggagg ctccggggag gggcgcgcg cggcgcgagg gccggcggt 1320
 gtcgagggcg gcgagccgc agccattgcc ttttatggta atcgtagcagg agggcgagg 1380
 gacttcctt gtcccaaatc tggcggagcc gaaatctggg aggcgcggcc gcacccctc 1440
 tagcgggcgc gggcgagcg ggcgggagcc ggcaggaagg aaatggggcg ggaggccct 1500
 cgtgcgtgc cgcgcgcgcg tccccttctc catctccagc ctccgggctg ccgaggggg 1560
 acggtgcct tcggggggga cggggcaggg cgggggttcg cttctggcgt gtgaccggcg 1620
 gctctagag ctctgctaac catgttcatt cctctctctt tttcttacag ctctggggc 1680
 acgtgctggt tgtgtgtgtg tctcatcatt ttggcaaaag attcactcct cagggtgcagg 1740
 ctgctatata gaaggtgggt gctgggtgtg ccaatgccct ggctcacaaa taccactgag 1800
 atctttttcc ctctgccaaa aattatgggg acatcatgaa gcccttgag catctgact 1860
 ctggctaata aaggaaattt attttcattg caatagtgtg ttggaatttt ttgtgtctct 1920
 cactcggaag gacatatggg agggcaaatc atttaaaaca tcagaatgag tatttggttt 1980
 agagtttggc aacatatgcc atatgctggc tgccatgaac aaaggtggct ataaagaggt 2040
 catcagtata tgaacagacc ccctgctgtc cattccttat tccatagaaa agccttgact 2100
 tgaggttaga ttttttttat attttgtttt gtgttatttt tttctttaac atccctaaaa 2160
 ttttccctac atgttttact agccagattt ttctctctct cctgactact ccagtcata 2220
 gctgtccctc ttctcttatg aagatccctc gacctgcagc ccaagcttgg cgtaatcatg 2280
 gtcatagctg tttcctgtgt gaaattgtta tccgctcaca attccacaca acatagcagc 2340
 cggaagcata aagtgtaaag cctgggggtg ctaatgagtg agctaactca cattaattgc 2400
 gttgcgtca ctgcccgtt tccagtcggg aaacctgtcg tgccagcgga tccgcatctc 2460
 aattagtcag caaccatagc cccgccccca actccgccc aactccgccc 2520
 agttccgccc attctccgcc ccctgctgag attccagaag tagtgaggag gcttttttgg aggcctaggg 2580
 gccgctcgg cctctgagct tttattgagc ttataaatgg ttacaaataa agcaatagca 2640
 ttttcaaaaa agctaaactt gcattttttt cactgcattc tagttgtggt ttgtccaaa 2700
 tcacaaattt cacaataaaa gtctggatcc gctgcattaa tgaatcgccc aacgcgcggg 2760
 tcatcaatgt atcttatcat ggcgctcttc cgcttctctg ctcactgact cgctgcgtc 2820
 gagaggcggt ttgcgtattg gggctatcgg cgggtatcgg gcggtaatag gggtatccac 2880
 ggtcgttcgg ctgcggcgag gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa 2940
 agaatacagg gccgcgttgc tggcggtttt ccataggtct cgccccctg acgagcatca 3000
 ccgtaaaaaa cgcctcaagt agaggtggcg aaacccgaca ggactataaa gataccagg 3060
 caaaaatcga cgtcgaagtc tcgtgcgtcc accctgccgc ttaccggata 3120
 gtttcccccc ggaagctccc

cctgtccgccc	tttctccctt	cgggaagcgt	ggcgctttct	caatgctcac	gctgtaggta	3240
tctcagttcg	gtgtaggtag	ttcgctccaa	gctgggctgt	gtgcacgaac	cccccgttca	3300
gcccgcaccg	tgccgccttat	ccggtaacta	tctgtctgag	tccaacccgg	taagacacga	3360
cttatcgcca	ctggcagcag	ccactggtaa	caggattagc	agagcgaggt	atgtaggcgg	3420
tgctacagag	ttcttgaagt	gggtggcctaa	ctacggctac	actagaagga	cagtatttgg	3480
tatctgcgct	ctgctgaagc	cagttacctt	cggaaaaaga	gttggtagct	cttgatccgg	3540
caaacaaacc	accgctggta	gcggtgggtt	ttttgtttgc	aagcagcaga	ttacgcgcag	3600
aaaaaaagga	tctcaagaag	atcctttgat	cttttctacg	gggtctgacg	ctcagtgga	3660
cgaaaactca	cgtaaagggg	ttttgggtcat	gagattatca	aaaaggatct	tcacctagat	3720
ccttttaaat	taaaaatgaa	gttttaaatc	aatctaaagt	atatatgagt	aaacttggtc	3780
tgacagttac	caatgcttaa	tcagtgaggg	acctatctca	gcgatctgtc	tatttcggtc	3840
atccatagtt	gcctgactcc	ccgtcgtgta	gataactacg	atacgggagg	gcttaccatc	3900
tggccccagt	gctgcaatga	taccgcgaga	cccacgctca	ccggctccag	atttatcagc	3960
aataaaccag	ccagccggaa	gggcccagcg	cagaagtggg	cctgcaactt	tatccgcctc	4020
catccagttc	attaattggt	gccgggaagc	tagagtaagt	agttcgccag	ttaatatgtt	4080
gcgcaacggt	gttgccattg	ctacaggcat	cgtgggtgtc	cgctcgtcgt	ttggtagggc	4140
ttcattcagc	tccggttccc	aacgatcaag	gcgagttaca	tgatccccca	tggtgtgcaa	4200
aaaagcgggt	agctccttcg	gtcctccgat	cgltgtcaga	agtaagttgg	ccgcagtggt	4260
atcactcatg	gttatggcag	cactgcataa	ttctcttact	gtcatgccat	ccgtaagatg	4320
gagatccagt	tcgatgtaac	actcaaccac	gtcattctga	gaatagtgtg	tgccggcgacc	4380
gagttgctct	tgcccggcgt	caatacggga	taataccgcg	ccacatagca	gaactttaaa	4440
agtgtctcat	attggaaaac	gttcttcggg	gcgaaaactc	tcaaggatct	taccgctggt	4500
gagatccagt	tcgatgtaac	ccactcgtgc	acccaactga	tcttcagcat	cttttacttt	4560
caccagcgtt	tctgggtgag	caaaaacagg	aaggcaaaat	gccgcaaaaa	agggataaag	4620
ggcgacacgg	aaatgttgaa	tactcatact	cttccttttt	caatattatt	gaagcattta	4680
tcagggttat	tgttctcatg	cgggatacat	atttgaaatg	atttagaaaa	ataaacaat	4740
aggggttccg	cgcacatttc	cccgaagagt	gccacctg			4778

<210> 71
 <211> 5510
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pCXeGFP plasmid

<400> 71						
gtcgacattg	attattgact	agttattaat	agtaatcaat	tacgggggtca	ttagttcata	60
gcccataatat	ggagttccgc	gttacataac	ttacggtaaa	tggcccgcct	ggctgaccgc	120
ccaacgaccc	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	acgccaatag	180
ggactttcca	ttgacgtcaa	tgggtggact	attacggtta	aactgcccac	ttggcagtag	240
atcaagtgtg	tcataatgcca	agtacgcccc	ctattgacgt	caatgacggt	aaatggcccc	300
cctggcattg	tgcccagtac	atgaccttat	gggactttcc	tacttggcag	tacatctacg	360
tattagtcac	cgctattacc	atgggtcgag	gtgagcccca	cgttctgctt	cactctcccc	420
atctcccccc	cccaattttg	tatttatatta	tatttatatta	ttttttaatt	attttgtgca	480
gcgatggggg	cggggggggg	ggggggcgcg	gccaggcggg	gcggggcggg	gcgagggggc	540
gggcggggcg	aggcggagag	gtgcggcgcg	agccaatcag	agcggcgcg	tccgaaagtt	600
tccttttatg	gcgagggcgg	ggcggcgggc	gccctataaa	aagcgaagcg	cgcgggcggg	660
gggagtcgct	gcgttgccct	cgccccgtgc	cccgcctcgc	gcgcgcctcg	gccgcggccc	720
ccggctctga	ctgaccgcgt	tactcccaca	ggtagcgggg	cgggacgggc	cttctcctcc	780
gggctgtaat	tagcgtctgg	tttaatgacg	gctcgtttct	tttctgtggc	tgctgtgaaag	840
ccttaaaggg	ctccgggagg	gcccttttgt	cgggggggag	cggtcggggg	ggtgcgtgcg	900
tgtgtgtgtg	cgtggggagc	gccgcgtgcg	gcccgcgctg	cccggcggct	gtgagcgctg	960
cgggcgcggc	gcggggcgtt	gtgcgctccg	cgtgtgcgcg	aggggagcgc	ggccgggggg	1020
ggtgccccgc	ggtgcggggg	ggctgcgagg	ggaacaaagg	ctgcgtgagg	ggtgtgtgcg	1080
tgggggggtg	agcagggggg	gtgggcgcgg	cggtcgggct	gtaacccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cggcccggtc	tcgggtgcgg	ggctccgtgc	ggggcggtgg	1200
gcggggctcg	ccgtgccggg	cgggggggtg	cggcaggttg	gggtgcgggg	cggggcgggg	1260
ccgcctcggg	ccggggaggg	ctcgggggag	gggcgcggcg	gccccggagg	gccggcggtc	1320
gtcagggcgc	gcgagccgc	agccattgcc	ttttatggta	atcgtagcag	agggcgaggg	1380
gacttccttt	gtcccaaatc	tggcggagcc	gaaatctggg	agggcgccgc	gcacccctcc	1440
tagcggggcg	gggcgaagcg	gtgcggcgcc	ggcaggaagg	aaatggggcg	ggagggcgct	1500
cgtgcgtcgc	cgcgcgcgcg	tccccttctc	catctccagg	ctcggggctg	cgcggggggg	1560
acggctgcct	tccggggggg	cggggcgagg	cgggggttcg	cttctggcgt	gtgacggggc	1620
gctctagagg	ctctgtctaac	catgttcatt	ccttcttctt	tttctctacg	ctcctgggca	1680
acgtgctggt	tgtgtgctg	tctcatcatt	ttggcaaaag	attcgccacc	atgggtgagca	1740
agggcgagga	gctgttcacc	gggggtggtg	ccatcctggt	cgagctggac	ggcgacgtaa	1800

acggccacaa	gttcagcgtg	tccggcgagg	gcgagggcga	tgccacctac	ggcaagctga	1860
ccctgaagtt	catctgcacc	acgggcaagc	tgcccgtgcc	ctggccacc	ctcgtgacca	1920
ccctgacctt	cgccgtgcag	tgcttcagcc	gtacccccga	ccacatgaag	cagcacgact	1980
tcttcaagtc	cgccatgccc	gaaggctacg	tccaggagcg	caccatcttc	ttcaaggacg	2040
acggcaacta	caagaccgcg	gcccagggtga	agttcgaggg	cgacaccttg	gtgaaccgca	2100
tccgagctgaa	gggcatcgac	ttcaaggagg	acggcaacat	cctggggcac	aagctggagt	2160
acaactacaa	cagccacaa	gtctatatca	tggccgacaa	gcagaagaac	ggcatcaagg	2220
tgaacttcaa	gatccgccac	aacatcgagg	acggcagcgt	gcagctcgcc	gaccactacc	2280
agcagaacac	ccccatcggc	gacggccccg	tgctgctgcc	cgacaaccac	tacctgagca	2340
cccagtcggc	cctgagcaaa	gacccccacg	agaagcgcg	tcacatggtc	ctgctggagt	2400
tccgtgaccgc	cgccggggtc	actctcgga	tggacagcgt	gtacaagtaa	gaattcactc	2460
ctcaggtgca	ggctgcctat	cagaagggtg	tggctgggtg	ggccaatgcc	ctggctcaca	2520
aataccactg	agatcttttt	ccctctgcca	aaaatttatg	ggacatcatg	aagccccctg	2580
agcattctgac	ttctggctaa	taaaggaaat	ttattttcat	tgcaatagtg	tggtgggaatt	2640
ttttgtgtct	ctcactcgga	aggacatatg	ggagggcaaa	tcatttaaaa	catcagaatg	2700
agtatttggg	ttagagtttg	gcaacatatg	ccatatgctg	gctgccatga	acaaagggtg	2760
ctataaagag	gtcatcagta	tatgaacag	ccccctgctg	tccattcctt	attccataga	2820
aaagccttga	cttgagggtta	gatttttttt	atattttgtt	ttgtgttatt	tttttcttta	2880
acatccctaa	aattttcctt	acatgtttta	ctagccagat	ttttcctcct	ctcctgacta	2940
ctcccgaatt	tagctgtccc	tcttctctta	tgaagatccc	tccagctgca	gccccagctt	3000
ggcgtaatca	tggctcatagc	tgtttctctg	gtgaaattgt	tatccgctca	caattccaca	3060
caacatacga	gcccgaagca	taaagtgtaa	agcctggggg	gcctaataag	tgagctaact	3120
cacattaatt	gcgttgcgct	cactgccgcg	tttccagtcg	ggaaacctgt	cgtgccagcg	3180
gatccgcate	tcaattagtc	agcaaccata	gtcccccccc	taactccgcc	catcccggcc	3240
ctaactccgc	ccagttccgc	ccatttctcg	ccccatgggt	gactaatttt	ttttatttat	3300
gcagaggccg	agggcgccct	ggcctctgag	ctattccaga	agtagtgagg	aggctttttt	3360
ggaggccctag	gcttttgcaa	aaagctaact	tgttttattg	agcttataat	ggttacaaat	3420
aaagcaatag	catcacaaat	ttcacaaata	aagcattttt	ttcactgcat	tctagttgtg	3480
gtttgtccaa	actcatcaat	gtatcttctc	atgtctggat	ccgctgcatt	aatgaatcgg	3540
ccaaecgcgc	gggagaggcg	gtttgcgtat	tgggcgctct	tccgcttctc	cgctcactga	3600
ctogetgcgc	tccggtcggtc	ggctgcggcg	agcgggtatc	gctcactcaa	aggcggtaat	3660
acggttatcc	acagaatcag	gggataacgc	aggaaagaac	atgtgagcaa	aaggccagca	3720
aaaggccatg	caacgttaaaa	aggccgcggt	gctggcggtt	ttccataggc	tccgcccccc	3780
tgacgagcat	cacaaaaatc	gacgctcaag	tccagaggtg	cgaaacccga	caggactata	3840
aagataccag	gcgtttcccc	ctggaagctc	cctcglgcgc	tctcctgttc	cgaccctgcc	3900
gcttaccgga	tacctgtccg	cctttctccc	ttcgggaagc	gtggcgcttt	ctcaatgctc	3960
acgctgtagg	tatctcagtt	cgggtgtaggt	cgttcgctcc	aagctgggct	gtgtgcacga	4020
accccccggt	cagcccgacc	gctgcgcctt	atccggtaac	tatcgtcttg	agtccaaccc	4080
ggtaagacac	cacttatcgc	cactggcagc	agccactggg	aacaggatta	gcagagcgag	4140
gtatgtaggc	ggtgctacag	agttcttgaa	gtgggtggcct	aactacggct	acactagaag	4200
gacagtatct	ggtatctgcg	ctctgctgaa	gccagttacc	ttcggaaaaa	gagttggtag	4260
ctcttgatcc	ggcaaacaaa	ccaccgctgg	tacgggtggg	ttttttgttt	gcaagcagca	4320
gattacgcgc	agaaaaaaag	gatctcaaga	agatcctttg	atcttttcta	cgggggtctga	4380
cgctcagtg	aacgaaaaat	cacgttaagg	gatttttggt	atgagattat	caaaaaaggat	4440
cttcacctag	atccttttaa	attaaaaatg	aagttttaaa	tcaatctaaa	gtatatatga	4500
gtaaacttgg	tctgacagtt	accaatgctt	aatcagtgag	gcacctatct	cagcgatctg	4560
tctatttctg	tcatccatag	ttgcctgact	ccccgtcgtg	tagataacta	cgatacggga	4620
gggcttacca	tctggcccca	gtgctgcaat	gataccgcga	gacccacgct	caccggctcc	4680
agatttatca	gcaataaacc	agccagccgg	aaggggccgag	cgcagaagtg	gtcctgcaac	4740
tttatccgcc	tccatccagt	ctattaattg	ttgcccggaa	gctagagtaa	gtagttcgcc	4800
agttaatagt	ttgcgcgaac	ttgttgccat	tgctacaggc	atcgtgggtg	cacgctcgtc	4860
gtttgggtatg	gcttcattca	gctccgggtc	ccaacgatca	aggcgagtta	catgatcccc	4920
catgttgtgc	aaaaaagcgg	ttagctcctt	cgggtcctccg	atcgttgtca	gaagtaagtt	4980
ggcgcgagtg	ttatcactca	tggttatggc	agcactgcat	aattctctta	ctgtcatgcc	5040
atccgtaaga	tgtttttctg	tgactgggtga	gtactcaacc	aagtcaattct	gagaatagtg	5100
tatgcggcga	cagagttgct	cttgcccggc	gtcaatacgg	gataataccg	cgccacatag	5160
cagaacttta	aaagtgtcca	tcattggaaa	acgttcttcg	gggcgaaaac	tctcaaggat	5220
cttaccgctg	ttgagatcca	gttcgatgta	accactcgt	gcacccaact	gatcttcagc	5280
atcttttact	ttcaccagcg	tttctgggtg	agcaaaaaaca	ggaaggcaaa	atgccgcaaa	5340
aaagggaata	agggcgacac	ggaaatgttg	aatactcata	ctcttctttt	ttcaatatta	5400
ttgaagcatt	tatcagggtt	attgtctcat	gagcgggatac	atatttgaat	gtatttagaa	5460
aaataaacia	ataggggttc	cgcgcacatt	tccccgaaaa	gtgccacctg		5510

<210> 72
 <211> 282
 <212> DNA
 <213> Artificial Sequence

<220>
<223> attp

<400> 72
ccttgcgcta atgctctgtt acaggctcact aataccatct aagtagttga ttcatagtga 60
ctgcatatgt tgtgttttac agtattatgt agtctgtttt ttatgcaaaa tctaatttaa 120
tatattgata tttatatcat tttacgtttc tcgttcagct tttttatact aagttggcat 180
tataaaaaag cattgcttat caatttggtg caacgaacag gtcactatca gtcaaaaataa 240
aatcattatt tgatttcaat tttgtccac tccctgcctc tg 282

<210> 73
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 73
ggccccgtaa tgcagaagaa 20

<210> 74
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 74
ggtttaagt gcgctcctcc aagaacgtca tc 32

<210> 75
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 75
agatctagag ccgccgtac aggaacaggt ggtggcgccc 40

<210> 76
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 5PacSV40

<400> 76
ctgttaatta actgtggaat gtgtgtcagt taggggtg 37

<210> 77
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer Antisense Zeo

<400> 77
tgaacagggt cacgtcgtcc 20

<210> 78
<211> 24

<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 5' HETS

<400> 78
gggccgaaac gatctcaacc tatt 24

<210> 79
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 3' HETS

<400> 79
cgcagcggcc ctcctactc 19

<210> 80
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 5BSD

<400> 80
accatgaaaa catttaacat ttctcaaca 29

<210> 81
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer SV40polyA

<400> 81
tttatttgtg aaatttgtga tgctattgc 29

<210> 82
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 3BSP

<400> 82
ttaatttcgg gtatatttga gtgga 25

<210> 83
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer EPO5XBA

<400> 83
tatctagaat ggggggtgcac gaatgtcctg cc 32

<210> 84
<211> 32
<212> DNA

<213> Artificial Sequence

<220>

<223> Primer EPO3SBI

<400> 84

tacgtacgtc atctgtcccc tgtcctgcag gc

32

<210> 85

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer GENEP03BSI

<400> 85

cgtagctcat ctgtccccctg tcctgca

27

<210> 86

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer GENEP05XBA

<400> 86

tctagaatgg ggggtgcacgg tgagtact

28

<210> 87

<211> 4862

<212> DNA

<213> Artificial Sequence

<220>

<223> pD2eGFP-1N plasmid from Clontech

<400> 87

tagttattaa tagtaatcaa ttacgggggtc attagttcat agcccatata tggagttccg 60
cgttacataaa cttacgggtaa atggccccgcc tggctgaccg cccaacgacc cccgcccatt 120
gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180
atgggtggag tatttacggg aaactgcccc cttggcagta catcaagtgt atcatatgcc 240
aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300
catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360
catggtgatg cgggttttggc agtacatcaa tgggcgtgga tagcgggttg actcacgggg 420
atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480
ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540
acgggtgggag gtctatataa gcagagctgg tttagtgaac cgtcagatcc gctagcgcta 600
ccggactcag atctcgagct caagcttcga attctgcagt cgacgggtacc gcgggcccgg 660
gatccaccgg tcgccaccat ggtgagcaag ggcgaggagc tggtcacccg ggtgggtgcc 720
atcctggtcg agctggacgg cgacgtaaac ggccacaagt tcagcgtgtc cggcgagggc 780
gagggcgatg ccacctacgg caagctgacc ctgaagttca tctgcaccac cggcaagctg 840
cccgtgccct ggcccacct cgtgaccacc ctgacctacg gcgtgcagtg cttcagccgc 900
taccocgacc acatgaagca gcacgacttc ttcaagtcgg ccatgcccga aggctacgtc 960
caggagcgca ccactcttct caaggacgac ggcaactaca agacccgcgc cgagggtgaag 1020
ttcgagggcg acacctgggt gaaccgcatc gagctgaagg gcatcgactt caaggaggac 1080
ggcaacatcc tggggcacaa gctggagtac aactacaaca gccacaacgt ctatatcatg 1140
gccgacaagc agaagaacgg catcaagggtg aacttcaaga tccgccacaa catcgaggac 1200
ggcagcgtgc agctcgccga ccactaccag cagaacaccc ccatcggcga cggccccgtg 1260
ctgctgcccc acaaccacta cctgagcacc cagtcgcgcc tgagcaaaga ccccaacgag 1320
aagcgcgatc acatggctct gctggagtgc gtgaccgcgg ccgggatcac tctcggcatg 1380
gacgagctgt acaagaagct tagccatggc ttcccgcggg aggtggagga gcaggatgat 1440
ggcagcgtgc ccatgtcttg tgcccaggag agcgggatgg accgtcacc tgcagcctgt 1500
gcttctgcta ggatcaatgt gtagatgcgc ggcgcgcagt ctagatcata atcagccata 1560
ccacatttgt agaggtttta cttgctttta aaaacctccc acacctccc ctgaacctga 1620
aacataaaat gaatgcaatt gttgttgtta acttgtttat tgcagcttat aatgggttaca 1680

```

aataaagcaa tagcatcaca aatttcacaa ataaagcatt tttttcactg cattctagtt 1740
gtggtttgtc caaactcatc aatgtatctt aaggcgtaaa ttgtaagcgt taatattttg 1800
ttaaaattcg cggttaaattt ttgttaaattc agctcattttt ttaaccaata ggccgaaatc 1860
ggcaaaatcc cttataaatc aaaagaatag accgagatag ggttgagtgt tggtccagtt 1920
tggaacaaga gtccactatt aaagaacgtg gactccaacg tcaaagggcg aaaaaccgtc 1980
tatcagggcg atggccact acgtgaacca tcaccctaata caagtttttt ggggtcgagg 2040
tgccgtaaag cactaaatcg gaaccctaaa gggagccccc gatttagagc ttgacgggga 2100
aagccggcga acgtggcgag aaagggaagg aagaaaagcga aaggagcggg cgctagggcg 2160
ctggcaagtg tagcggtcac gctgcgcgta accaccacac ccgcccgcgt taatgcgcgc 2220
ctacagggcg cgtcaggtgg cacttttcgg ggaaatgtgc gcggaacccc tatttgttta 2280
tttttctaaa tacattcaaa tatgtatccg ctcatgagac aataaccctg ataaatgctt 2340
caataatatt gaaaaaggaa gagtcctgag gcggaagaa ccagctgtgg aatgtgtgtc 2400
agttagggtg tggaaagtcc ccaggctccc cagcaggcag aagtatgcaa agcatgcatc 2460
tcaatttagc agcaaccagg tgtggaagt cccaggctc cccagcaggc agaagatgc 2520
aaagcatgca tctcaattag tctgcaacca tagtcccgcc cctaactccg cccatccgcg 2580
ccctaactcc gccaggttcc gccattctc cgccccatgg ctgactaatt ttttttattt 2640
atgcagggcg agggcttttc tggcctctg agctattcca gaagtagtga ggaggctttt 2700
ttggaggcct aggtttttgc aaagatcgat caagagacag gatgaggatc gtttcgcatg 2760
attgaacaag atggattgca cgcaggttct ccggccgctt gatggagag gctattcgcg 2820
tatgactggg cacaacagac aatcggctgc tctgatgccg cctgtgtccg gctgtcagcg 2880
cagggcgccg cggttctttt tgtcaagacc gacctgtccg gtgccctgaa tgaactgcaa 2940
gacgaggcag cgcggtatc gtggctggcc acgacggggc gttccttgcc agctgtgtct 3000
gacgttgtca ctgaagcggg aagggaactg ctgctattgg cgaagtgcc ggggcaggat 3060
ctcctgtcat ctcaccttgc tcctgccgag aaagtatcca tcatggctga tgcaatgcgg 3120
cggtgtcata cgcttgatcc ggctacctgc ccattcgacc accaagcgaa acatcgcatc 3180
gagcagacac gtactcggat ggaagccggg cttgtcgatc aggatgatc ggacgaagag 3240
catcaggggc tcgcgccagc cgaactgttc cccaggtca aggcgagcat gccgcaggc 3300
gaggatctcg tegtgaacca tggcgatgcc tgcttgccga atatcatggt ggaaaatgac 3360
cgcttttctg gattcatcga ctgtggccgg ctgggtgtgg cggaccgcta tcaggacata 3420
gcgttggtca ccctgatata tgctgaagag cttggcgccg aatgggctga ccgttctctc 3480
gtgctttacg gtatcgccgc tcccgatctg cagcgcatcg ccttctatcg ccttcttgac 3540
gagttcttct gagcgggact ctgggggttcg aaatgaccga ccaagcgacg cccaacctgc 3600
catcacgaga tttcgattcc accgccgctt tctatgaaag gttgggcttc ggaatcgttt 3660
tccgggacgc cggctggatg atcctccagc gctgtgaaag gttgggcttc catgctggag 3720
accctagggg gaggctaact gaaacacgga aggagacaat accggaagga acccgcgcta 3780
tgacggcaat aaaaaacgc aataaaacgc acggtgttgg gtcgtttggt cataaacgcg 3840
gggttcgggt ccagggctgg cactctgtcg atacccacc gagaccccat tggggccaat 3900
acgcccgcgt tcttctcttt cccccacccc ccccccaag ttccgggtgaa ggcccagggc 3960
tcgcagccaa cgtcggggcg agtgcctctg ccttgccctc aggttactca tatatacttt 4020
agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata 4080
atctcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccccgtag 4140
aaaaatcctt aggatcctt tgagatcctt cttggcgccg cgtaatctgc tgcttgcaa 4200
caaaaaaacc accgctacca gcggtggttt gtttgccgga tcaagagcta ccaactcttt 4260
ttccgaaggt aactggcttc agcagagcgc agataccaaa tactgtcctt ctagtgtagc 4320
cgtagttagg ccaccacttc aagaactctg tagcacgcgc gctctgctaa gctctgctaa 4380
tctgtttacc agtggtgct gccagtggcg gccagcgcgc tacatacctc ccaactcttt 4440
gacgatagtt accggataag gcgcagcggg acaggtatcc gggggggttcg tgccacacagc 4500
ccagcttgga gcgaacgacc tacaccgaac acagcgtagg ctatgagaaa 4560
gcgccacgct tcccgaaggg agaaaggcgg agaggtatcc ggtaagcggc agggctcgga 4620
caggagagcg caccagggag cttccagggg gaaacgcctg gtatctttat agtcctgtcg 4680
ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg gggcgaggcc 4740
tatggaaaaa cgccagcaac gcggcctttt tacggttcct ggccttttgc tggccttttg 4800
ctcacatggt ctttctgcg ttatccctcg attctgtgga taaccgtatt accgccatgc 4860
at

```

<210> 88
 <211> 5192
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pIRESpuro2 plasmid from Clontech

<400> 88
 gacggatcgg gagatctccc gatcccttat ggtcgactct cagtacaatc tgctctgatg 60
 ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gactagtgcg 120
 cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc 180

ttaggggttag	gcgtttttgcg	ctgctttcgcg	atgtacgggc	cagatatagc	cggtgacatt	240
gattatttgac	tagttatttaa	tagtaatacaa	ttacgggggtc	attagttcat	agcccatata	300
tggaggttccg	cgttacataaa	cttacgggtaa	atggcccgcc	tggctgaccg	cccaacgacc	360
cccggccatt	gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	420
attgacgtca	atgggtgggac	tattttacggg	aaactgcccc	cttgggcagta	catcaagtgt	480
atcatalgccc	aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	540
atggcccagta	catgacctta	tgggactttc	ctacttgcca	gtacatctac	gtattagtca	600
tcgctatttac	catgttgatg	cggtttttggc	agtacatcaa	tgggcggtgga	tagcgggtttg	660
actcacgggg	atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	720
aaaatcaacg	ggactttcca	aaatgtcgta	acaactccgc	cccattgacg	caaatgggcg	780
gtaggcggtgt	acgggtgggag	gtctatataaa	gcagagctct	ctgggctaact	agagaaccca	840
ctgcttactg	gcttatcgaa	attaatacga	ctcactatag	ggagacccaa	gcttggtacc	900
gagctcggat	cgatatctgc	ggcctagcta	gcgcttaagg	cctgttaacc	ggtcgtagct	960
ctccggattc	gaattcggat	ccgcgggcgc	atagataaact	gatccagtgt	gctggaatta	1020
attcgctgtc	tgcgagggcc	agctgttggg	gtgagtactc	cctctcaaaa	gcgggcatga	1080
cttctgcgct	aagattgtca	gtttccaaaa	acgaggagga	tttgatattc	acctggcccc	1140
cggtgatgcc	tttgagggtg	gccgcgtcca	ctgggtcaga	aaagacaatc	ttttgtttgt	1200
caagcttgag	gtgtggcagg	cttgagatct	ggccatacac	ttgagtgaca	atgacatcca	1260
ctttgccttt	ctctccacag	gtgtccactc	ccaggtccaa	ctgcaggctcg	agcatgcatc	1320
tagggcgggc	aattccgccc	ctctccctcc	cccccccta	acgttactgg	ccgaacccca	1380
ttggaataag	gccggtgtgc	gtttgtctat	atgtgatttt	ccaccatatt	gccgtctttt	1440
gggcaatgtga	ggggccgggaa	acctggccct	gtcttcttga	cgagcatctc	taggggtctt	1500
tccctctcgc	ccaaaggaat	gcaaggctcg	ttgaatgtcg	tgaaggaagc	agttccctctg	1560
gaagcttctt	gaagacaaac	aacgtctgta	gcgacccttt	gcaggcagcg	gaacccccca	1620
cctggcgaca	ggtgcctctg	cggccaaaaag	ccagctgtat	aagatacacc	tgcaaacggcg	1680
gcacaacccc	agtgccacgt	tgtgagttgg	atagttgtgg	aaagagtcaa	atggctcttc	1740
tcaagcgtat	tcaacaaggg	gctgaaggat	gcccagaagg	taccccatgg	tatgggatct	1800
gatctggggc	ctcggtgcac	atgctttaca	tgtgtttagt	cgagggttaa	aaaacgtcta	1860
ggccccccga	accacgggga	cgtggttttc	ctttgaaaaa	cacgatgata	agcttgccac	1920
aaaccacaa	gagacgacct	tccatgaccg	agtacaaagc	cacgggtgcg	ctcgccaccc	1980
gcgacgacgt	cccccgggcc	gtacgcaccc	tcgcccggcg	gttcgcccgc	taccccgcca	2040
cgcgccacac	cgtcgacccg	gaccgcaca	tcgagcgggt	caccgagctg	caagaactct	2100
tcctcacgcg	cgtcgggttc	gacatcgcca	agggtgtggg	cgcgagcgac	ggcgcccgcg	2160
tggcggtctg	gaccacgccc	gagagcgtcg	aagcgggggc	ggtgttcgcc	gagatcgccc	2220
cgcgcatggc	cgagttgagc	ggttcccggc	tggccgcgca	gcaacagatg	gaaggcctcc	2280
tggcgccgca	ccggcccaag	gagcccgcgt	ggttccctgg	caccgtcggc	gtctcgcccc	2340
accaccaggg	caaggggtctg	ggcagcgccg	tcgtgtctcc	cgaggtggag	gcggccgagc	2400
gcgcgggggt	gcccgccttc	ctggagacct	cccgcccccg	caacctcccc	ttctacgagc	2460
ggctcggctt	caccgtcacc	gcccgcgtcg	agtgcccga	ggaccgcgcy	acctggtgca	2520
tgaccgcgaa	gcccgggtgcc	tgacgcccgc	cccacgaccc	gcagcgcccc	accgaaaagg	2580
gcgcacgacc	ccatggctcc	gaccgaagcc	gacccgggcg	gccccgcgca	ccccgcaccc	2640
cggggaaacct	cccaccgact	ctagagctcg	ctgatcagcc	tcgactgtgc	cttctagtgt	2700
ccagccatct	gttgtttgccc	cctccccctg	gccttccctg	accctgggaag	gtgccactcc	2760
cactgtcctt	tcctaataaa	atgaggaaat	tgcatcgcat	tgtctgagta	ggtgtcattc	2820
tattctgggg	ggtgggggtg	ggcaggacag	caagggggag	gattggggaag	acaatagcag	2880
gcatgctggg	gatgcgggtg	gctctatggc	ttctgaggcg	gaaagaacca	gctggggctc	2940
gagtgcattc	tagttgtggg	ttgtccaaac	tcataaatgt	atcttatcat	gtctgtatac	3000
cgtcgacctc	tagctagagc	ttggcgtaat	catggtcata	gctgttttct	gtgtgaaatt	3060
gttatccgct	cacaattcca	cacaacatac	gagccggaag	cataaagtgt	aaagcctggg	3120
gtgcctaata	agtgaagctaa	ctcacattaa	ttgcgttgcc	ctcactgccc	gctttccagt	3180
cggggaaacct	gtcgtgccag	ctgcattaat	gaatcggccca	acgcgcgggg	agaggcggtt	3240
tgcgtattgg	gcgctcttcc	gcttccctcg	tcactgactc	gctgcgctcg	gtcgttcggc	3300
tgcggcgagc	ggtatcagct	cactcaaaag	cggtataacg	ggttatccaca	gaatcagggg	3360
ataacgcagg	aaagaacatg	tgagcaaaaag	gccagcaaaa	ggccagggaac	cgtaaaaagg	3420
ccgcgttgct	ggcggttttt	cataggctcc	gccccctga	cgagcatcac	aaaaatcgac	3480
gctcaagtca	gaggtggcga	aacccgacag	gactataaag	ataccaggcg	tttccccctg	3540
gaagctccct	cgtgcgctct	cctgttccga	ccctgccgct	taccggatac	ctgtccgcct	3600
ttctcccttc	gggaagcgtg	gcgctttctc	aatgctcacg	ctgtagggtat	ctcagttcgg	3660
tgtaggctgt	tcgctccaag	ctgggctgtg	tgacgaacc	ccccgttcag	cccgaccgct	3720
gcgccttatc	cgtcttgagt	cgtcttgagt	ccaaccgggt	aagacacgac	ttatcgccac	3780
tggcagcagc	cactggtaac	aggattagca	gagcgaggta	tgtaggcggt	gctacagagt	3840
tcttgaagtg	gtggccctaac	tacggctaca	ctagaaggac	agtatttggg	atctgcgctc	3900
tgctgaagcc	agttaccttc	ggaaaaagag	ttggtagctc	ttgatccggc	aaacaaaagg	3960
ccgctggtag	cggtgggttt	tttgtttgca	agcagcagat	tacgcgcaga	aaaaaaggat	4020
ctcaagaaga	tccctttgatc	ttttctacgg	ggtctgacgc	tcagtgggaac	gaaaactcac	4080
gttaagggat	tttggctatg	agattatcaa	aaaggatctt	cacctagatc	cttttaaaatt	4140
aaaaatgaag	ttttaaatca	atctaaagta	tatatgagta	aacttgggtct	gacagttacc	4200

aatgcttaaat	cagttaggca	cctatctcag	cgatctgtct	atttcgttca	tccatagttg	4260
cctgactccc	cgctcgttag	ataactacga	tacgggaggg	cttaccatct	ggccccagtg	4320
ctgcaatgat	accgcgagac	ccacgctcac	cggctccaga	tttatcagca	ataaaccagc	4380
cagccggaag	ggccgagcgc	agaagtggtc	ctgcaacttt	atccgcctcc	atccagtcta	4440
ttaattgttg	ccgggaagct	agagtaagta	gttcgccagt	taatagtttg	cgcaacgttg	4500
ttgccattgc	tacaggcatc	gtggtgtcac	gctcgtcggt	tggatggct	tcattcagct	4560
ccggttccca	acgatcaagg	cgagttacat	gatcccccat	gttgtgcaaa	aaagcgggta	4620
gctccttcgg	tcctccgata	gttgtcagaa	gtaagtggc	cgagtggtta	tcactcatgg	4680
ttatggcagc	actgcataat	tctcttactg	tcatgccatc	cgtaagatgc	ttttctgtga	4740
ctggtgagta	ctcaaccaag	tcattctgag	aatagtgtat	gcggcgaccg	agttgtctct	4800
gcccggcgct	aatacgggat	aataccgcgc	cacatagcag	aactttaaaa	gtgctcatca	4860
ttggaaaaacg	ttcttcgggg	cgaaaaactct	caaggatctt	accgctgttg	agatccagtt	4920
cgatgtaacc	cactcgtgca	cccaactgat	cttcagcatc	ttttactttc	accagcggtt	4980
ctgggtgagc	aaaaacagga	aggcaaaaata	ccgcataaaa	gggaataagg	gcgacacgga	5040
aatgttgaat	actcatactc	ttcctttttc	aatattattg	aagcatttat	caggggttatt	5100
gtctcatgag	cggatacata	tttgaatgta	tttagaaaaa	taaacaaaata	gggggtccgc	5160
gcacattttcc	ccgaaaagtg	ccacctgacg	tc			5192

<210> 89
 <211> 11182
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pAg1 Plasmid

<400> 89						
catgccaaacc	acaggggttcc	cctcggggtc	aaagtacttt	gatccaaccc	ctccgctgct	60
atagtgcagt	cggcttctga	cggtcagtcg	agccgtcttc	tgaaaacgac	atgtcgcaca	120
agtcctaagt	tacgcgacag	gctgccgccc	tgcccttttc	ctggcgtttt	cttgtcgcgt	180
gtttttagtcg	cataaagttag	aataacttgcg	actagaaccg	gagacattac	gccatgaaca	240
agagcgccgc	cgctggcctg	ctgggctatg	ccgcgctcag	caccgacgac	caggacttga	300
ccaaccaaacg	ggccgaactg	cacgcggcgc	gctgcaccaa	gctgttttcc	gagaagatca	360
ccggcaccag	gcgcgaccgc	ccggagctgg	ccaggatgct	tgaccacctt	cgccctggcg	420
acgttgtgac	agtgaccagg	ctagaccgcc	tggcccgag	cacccgcgac	ctactggaca	480
ttgccgagcg	catccaggag	gccggcgccg	gcctgcgtag	cctggcagag	ccgtggggcg	540
acaccaccac	gccggccggc	cgcatgggtg	tgaccgtgtt	cgccggcatt	gccgagttcg	600
agcgttccct	aatcatcgac	cgcaccgcga	gcgggcgcga	ggccgcgaag	gcccagggcg	660
tgaagtttgg	ccccgccttc	accctcaccc	cgccacagat	cgcgacgcc	cgcgagctga	720
tcgaccagga	aggccgcacc	gtgaaagagg	cggctgcact	gcttggcggt	catcgctcga	780
ccctgtaccg	cgcacttgag	cgcagcgagg	aagtgcgcgc	caccgaggcc	aggcggcgcg	840
gtgccttccg	tgaggacgca	ttgaccgagg	ccgacgccct	ggcgcccgcc	gagaatgaac	900
gccaagagga	acaagcatga	aaccgcacca	ggacggccag	gacgaaccgt	ttttcattac	960
cgaagagatc	gaggcggaga	tgatcgcggc	cgggtacgtg	ttcgagccgc	ccgcgcacgt	1020
ctcaaccgtg	cggctgcatg	aaatcctggc	cggtttgtct	gatgccaaag	tggcgccctg	1080
gccggccagc	ttggccgctg	aagaaaaccga	gcgcgcgcgt	ctaaaaaggt	gatgtgtatt	1140
tgagtaaaac	agcttgcgct	atgcggtcgc	tgcgtatatg	atgcgatgag	taaataaaca	1200
aatacgcaag	gggaacgcac	gaagggtatc	gctgtactta	accagaaagg	cgggtcaggg	1260
aagacgacca	tcgcaaccca	tctagcccg	gccctgcaac	tcgcccgggg	cgatgttctg	1320
ttagtgcatt	ccgatcccca	gggcagtgcc	cgcgattggg	cgcccggtgc	ggaagatcaa	1380
ccgctaaccg	ttgtcggcat	cgaccggccc	acgattgacc	gcgacgtgaa	ggccatcggc	1440
cggcgcgact	tcgtagtgat	cgacggagcg	ccccaggcgg	cggacttggc	tgtgtccgcg	1500
atcaaggcag	ccgacttcgt	gctgattccg	gtgcagccaa	gcccttacga	catatgggcc	1560
accgccgacc	tggtggagct	gggttaagcag	cgcattgagg	tcacggatgg	aaggctacaa	1620
gccggcctttg	tcgtgtcgcg	ggcgatcaaa	ggcacgcgca	tcggcggtga	ggttgccgag	1680
gcgctggccg	ggtagcagct	gcccattctt	gagtcgccga	tcacgcagcg	cgtgagctac	1740
ccaggcactg	ccgcgcgcgg	cacaaccggt	cttgaatcag	aacccgaggg	cgacgctgcc	1800
cgcgagggtcc	aggcgctggc	cgctgaaatt	aaatcaaaac	tcatttgagt	taatgaggtg	1860
aagagaaaaa	gagcaaaaag	acaaaacacgc	taagtgcggg	ccgtccgagc	gcacgcagca	1920
gcaaggctgc	aacgtttggc	agcctggcag	acacgcagc	catgaagcgg	gtcaactttc	1980
agttgcccgg	ggaggatcac	accaagctga	agatgtacgc	ggtacgccaa	ggcaagacca	2040
ttaccgagct	gctatctgaa	tacatcgcg	agctaccaga	gtaaatgagc	aaatgaataa	2100
atgagttagt	gaatttttag	ggctaaaagg	ggcggcagtg	aaaatcaaga	acaaccaggc	2160
accgacgcgg	tggaaatgcc	catgtgtgga	ggaacgggcg	gttggccagg	cgtaagcggc	2220
tgggttgtct	gccggccctg	caatggcact	ggaaccccca	agcccagagg	atcggcggtg	2280
cggtcgcaaa	ccatccggcc	cggtaaaaaa	cggcgcgcg	ctgggtgatg	acctggtgga	2340
gaagttgaag	gccgcgcagg	ccgcccagcg	gcaacgcac	gaggcagaag	cacgccccgg	2400

tgaatcgtgg	caagcggccg	ctgatcgaat	ccgcaaagaa	tcccggcaac	cgccggcagc	2460
cggtgcgccg	tccgtaggga	agccgccccaa	gggcgacgag	caaccagatt	ttttcgttcc	2520
gatgctctat	gacgtgggca	cccgcgatag	tccgagcatc	atggacgtgg	ccgttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	ggtgatccgc	tacgagcttc	cagacgggca	2640
cgtagagggt	tccgcagggc	cggccggcat	ggccagtggt	tgggattacg	acctgggtact	2700
gatggcgggt	tcccatctaa	ccgaatccat	gaaccgatac	cgggaaggga	agggagacaa	2760
gcccggccgc	gtgttccgct	cacacgttgc	ggacgtactc	aagttctgcc	ggcgagccga	2820
tggcggaaag	cagaaagacg	acctggtaga	aaaatggcatt	cggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cggccgcctg	gtgacgggtat	ccgaggggtga	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaaaacc	gggcccggcg	agtacatcga	3000
gatcgagcta	gctgattgga	tgtaccgcga	gatcacagaa	ggcaagaacc	cggacgtgct	3060
gacgggttcac	cccgattact	ttttgatcga	tcccggcatc	ggccggtttc	tctaccgcct	3120
ggcaccggcg	gcccgcaggca	aggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
cagtggcagc	gcccgaaggt	tcaagaagtt	ctggttcacc	gtgcgcaagc	tgatcgggtc	3240
aaatgacctg	ccggagtacg	atttgaagga	ggaggcgggg	caggctggcc	cgatcctagt	3300
catgcgctac	cgcaaacctga	tccgagggcg	agcatccgcc	ggttcctaat	gtacggagca	3360
gatcgtaggg	caaattgccc	tagcagggcg	aaaaggctga	aaaggtctct	ttcctgtgga	3420
tagcacgtac	attgggaacc	caaagccgta	cattgggaac	cggaacccgt	acattgggaa	3480
cccaaaagccg	tacattggga	accgggtcaca	catgtaagtg	actgatataa	aagagaaaaa	3540
aggcgatttt	tcgcgcctaaa	actctttaa	acttattaaa	actctttaa	cccgcctggc	3600
ctgtgcataa	ctgtctggcc	agcgcacagc	cgaagagctg	caaaaagcgc	ctacccttcg	3660
gtcgtgcgcg	tccctacgcc	ccgcgcgttc	gcgtgcgcct	atcgccggcg	ctggccgctc	3720
aaaaatggct	ggcctacggc	caggcaatct	accagggcgc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	cggcgcccac	atcaaggcac	cctgcctcgc	gcgtttcggt	gatgacgggtg	3840
aaaacctctg	acacatgcag	ctcccggaga	cggtcacagc	ttgtctgtaa	gcggatgccg	3900
ggagcagaca	agcccgtcag	ggcgcgctcag	cgggtgttgg	cgggtgtcgg	ggcgagcca	3960
tgaccagtc	acgtagcgat	agcggagtgt	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgcacc	atatggcggtg	tgaaataaccg	cacagatgcg	taaggagaaa	4080
ataccgcatc	aggcgctctt	ccgcttcctc	gctcactgac	tcgctgcgct	cggtcggtcg	4140
gctgcggcga	gcgggtatcag	ctcactcaaa	ggcggttaata	cggttatcca	cagaatcagg	4200
ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	4260
ggccgcgttg	ctggcggtttt	tccataggct	ccgccccct	gacgagcatc	acaaaaatcg	4320
acgctcaagt	cagaggtggc	gaaacccgac	aggactataa	agataccagg	cgtttccccc	4380
tggaagctcc	ctcgtgcgct	ctcctgttcc	gaccctgcgc	cttaccggat	acctgtccgc	4440
ctttctccct	tcgggaagcg	tggcgctttc	tcatagctca	cgctgtaggt	atctcagttc	4500
ggtgtagggtc	gttcgctcca	agctgggctg	tgtgcacgaa	ccccccgttc	agcccgaaccg	4560
ctgcgcctta	tccggtaaact	atcgtcttga	gtccaacccg	gtaagacacg	acttatcgcc	4620
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	4680
gtctctgaag	tgggtggccta	actacggcta	actcagaagg	acagtatttg	gtatctgcgc	4740
tctgctgaag	ccagttaccct	tccgaaaaaag	agttggtagc	tcttgatccg	gcaaaacaaac	4800
caccgctgggt	agcgggtgggt	ttttttgtttg	caagcagcag	attacgcgca	gaaaaaaaagg	4860
atctcaagaa	gatcctttga	tctttctctac	ggggtctgac	gctcagtgga	acgaaaaactc	4920
acgttaagggt	atttttggtca	tgcattctag	gtactaaaaac	aattcatcca	gtaaaaatata	4980
atatttttatt	ttctcccaat	caggcttgat	ccccagtaag	tcaaaaaata	gctcgacata	5040
ctgttctctcc	ccgatatcct	ccgtgatcga	ccggagcgag	aaggcaatgt	cataccactt	5100
gtccgcctctg	ccgcttctcc	caagatcaat	aaagccactt	actttgccat	ctttcacaaa	5160
gatgttgctg	tctcccagggt	cgccgtggga	aaagacaagt	tcctcttcgg	gcttttccgt	5220
ctttaaaaaaa	tcatacagct	cgcgcggatc	ttttaaattgga	gtgtcttctt	cccagttttc	5280
gcaatccaca	tccggccagat	cgttattcag	taagtaaatcc	aattcggcta	agcggctgtc	5340
taagctattc	gtatagggac	aatccgatat	gtcgatggag	tgaaagagcc	tgatgcactc	5400
cgcatacagc	tcgataatct	tttcagggtc	ttgttcatct	tcatactctt	ccgagcaaaag	5460
gacgccatcg	gcctcactca	tgagcagatt	gctccagcca	tcattgcggt	caaagtgcag	5520
gaccttttga	acaggcagct	ttccttccag	ccatagcatc	atgtcctttt	cccgttccac	5580
atcatagggtg	gtccctttat	accggctgtc	cgctattttt	aaatatagggt	tttcattttc	5640
tcccaccagc	ttatatacct	tagcaggaga	cattccttcc	gtatctttta	cgccagcggt	5700
tttttcgatc	agtttttttca	attccgggtga	tattctcatt	ttagccattt	attatttctc	5760
tcctcttttc	tacagtattt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
aattcactgt	tccttgcat	ctaaaacctt	aaataccaga	aaacagcttt	ttcaaagtgt	5880
ttttcaaagt	tggcgataaa	catagtatcg	acggagccga	ttttgaaacc	gcgggtgatca	5940
caggcagcaa	cgtctgttca	tcgttacaat	caacatgccta	ccctccgcga	gatcatccgt	6000
gtttcaaacc	cggcagctta	gttgccgttc	ttccgaatag	catcggtaac	atgagcaaaag	6060
tctgccgcct	tacaacggct	ctcccgctga	cgcgcgtccg	gactgatggg	ctgcctgtat	6120
cgagtgggtga	ttttgtgccc	agctgcccgt	cggggagctg	ttggctgggt	ggtggcagga	6180
tatatgtgtg	tgtaaacaaa	ttgacgctta	gacaacttaa	taacacattg	cggacgtttt	6240
taattgtactg	aattaacgcc	gaattaatc	gggggatctg	gatttttagta	ctggattttt	6300
tttttaggaa	ttagaaattt	tattgataga	agtattttac	aaatacaaat	acatactaag	6360
ggtttcttat	atgctcaaca	catgagcgaa	accctatagg	aaccctaatt	cccttatctg	6420

ggaactactc	acacattatt	atggagaaac	tcgagtcaaa	tctcgggtgac	gggcaggacc	6480
ggacggggcg	gtaccggcag	gctgaagtcc	agctgccaga	aaccacagtc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgccctcgtc	6600
atgcgcacgc	tccgggtcgt	gggcagcccg	atgacagcga	ccacgctctt	gaagccctgt	6660
gcctccaggg	acttcagcag	gtgggtgtag	agcgtggagc	ccagtcocgt	ccgctgggtg	6720
cgggggggaga	cgtacacggg	cgactcggcc	gtccagtcgt	aggcgtttgc	tgccttccag	6780
ggggcccgct	aggcgatgcc	ggcgaccccg	ccgtccacct	cggcgacgag	ccagggatag	6840
cgctcccga	gacggacgag	gtcgtccgtc	cactcctgcg	gttcctgcgg	ctcggtagcg	6900
aagttgaccg	tgcttgtctc	gatgtagtgg	ttgacgatgg	tgcagaccgc	cggcatgtcc	6960
gcctcgggtg	cacggcggat	gtcggccggg	cgtcgttctg	ggctcatggg	agactcgaga	7020
gagatagatt	tgtagagaga	gactggtgat	ttcagcgtgt	cctctccaaa	tgaaatgaac	7080
ttccttatata	agaggaagg	cttgcgagg	atagtgggat	tgtgcgtcat	cccttacgtc	7140
agtggagata	tcacatcaat	ccacttgctt	tgaagacgtg	gttgggaacgt	cttctttttc	7200
cacgatgctc	ctcgtgggtg	ggggcccatc	tttgggacca	ctgtcggcag	aggcatcttc	7260
aacgatagcc	tttcccttat	cgcaatgatg	gcattttgtg	gtgccacctt	ctttttctac	7320
tgtccttttg	atgaagtgc	agatagctgg	gcaatggaat	ccgaggaggt	ttcccgatat	7380
taccctttgt	tgaaaagtct	caatagccct	ttggctctct	gagactgtat	ctttgatatt	7440
cttggagtag	acgagagtgt	cgtgctccac	catgttatca	catcaatcca	cttgctttga	7500
agacgtggtt	ggaacgtctt	ctttttccac	gatgctcctc	gtgggtgggg	gtccatcttt	7560
gggaccactg	tcggcagagg	catcttgaac	gatagccttt	cctttatcgc	aatgatggca	7620
tttgtaggtg	ccaccttctt	tttctactgt	ccttttgatg	aagtgcacga	tagctgggca	7680
atggaatccg	aggagggttc	ccgatattac	cctttgttga	aaagtctcaa	tagccctttg	7740
gtcctctgag	actgtatctt	tgatattctt	ggagtagacg	agagtgtcgt	gctccaccat	7800
gttggcaagc	tgctctagcc	aatacgcaaa	ccgcctctcc	ccgcgcgttg	gccgattcat	7860
taatgcagct	ggcacgcacg	gtttcccgac	tggaaagcgg	gcagtgagcg	caacgcaatt	7920
aatgtgagtt	agctcactca	ttaggcaccc	caggctttac	actttatgct	tcgggctcgt	7980
atgttgtgtg	gaattgtgag	cggaatacaa	tttcacacag	gaaacagcta	tgaccatgat	8040
tacgaattcg	agccttgact	agagggtcga	cggtatacag	acatgataag	atacattgat	8100
gagtttgagc	aaaccacaac	tagaatgcag	tgaaaaaaat	gctttatttg	tgaaatttgt	8160
gatgctattg	ctttatttgt	aaccattata	agctgcaata	aacaagttgg	ggtgggcaaa	8220
gaactccagc	atgagatccc	cgcgctggag	gatcatccag	ccggcgctcc	ggaaaaacgat	8280
tccgaagccc	aacctttcat	agaaggcggc	ggtggaatcg	aaatctcgta	gcacgtgtca	8340
gtcctgtccc	tcggccacga	agtgcacgca	gttgccggcc	gggtcgcgca	gggcgaactc	8400
ccgccccac	ggctgctcgc	cgatctcggt	catggccggc	ccggaggcgt	cccggaagtt	8460
cgtggacacg	acctccgacc	actcggcgta	cagctcgtcc	aggccgcgca	cccacaccca	8520
ggccagggtg	ttgtccggca	ccacctggtc	ctggaccggc	ctgatgaaca	gggtcacgtc	8580
gtcccggacc	acaccggcga	agtcgtcctc	cacgaagtc	cgggagaacc	cgagccgggtc	8640
ggtccagaac	tcgacgcgtc	cggcgcagctc	gcgcgcgggtg	agcaccggaa	cggcactggt	8700
caacttggcc	atgtggtccg	atttcgctca	agttagtata	aaaaagcagg	cttcaatcct	8760
gcagggaattc	gatcgacact	ctcgtctact	ccaagaatat	caaagataca	gtctcagaag	8820
accaaaaggg	tattgagact	tttcaacaaa	gggtaatatc	gggaaacctc	ctcggattcc	8880
cttcccagca	tatctgtcac	ttcatcaaaa	gggacagtga	aaaggaagg	ggcacctaca	8940
aatgccatca	ttgcgataaa	ggaaaggcta	tcgttcaaga	tgcctctgcc	gacagtgggtc	9000
ccaaaagtgg	acccccaccc	acgaggagca	tcgtggaaaa	agaagacgtt	ccaaccacgt	9060
cttcaaaagca	agtggtatga	tgtgataaca	tgggtggagca	cgacactctc	gtctactcca	9120
agaatatcaa	agatacagtc	tcagaagacc	aaagggtat	tgagactttt	caacaaaggg	9180
taatatcggg	aaacctcctc	ggattccatt	gcccagctat	ctgtcacttc	atcaaaagga	9240
cagtagaaaa	ggaagggtgg	acctacaaat	gccatcattg	cgataaagga	aaggctatcg	9300
ttcaagatgc	ctctgccgac	agtgggtcca	aagatggacc	cccacccacg	aggagcatcg	9360
tggaaaaaaga	agacgttcca	accacgtctt	caaagcaagt	ggattgatgt	gatattctcca	9420
ctgacgtaag	ggatgacgca	caatcccact	atccttcgca	agaccttctt	ctatataagg	9480
aagttcattt	catttgagga	ggacacgctg	aaatcaccag	tctctctcta	caaactctatc	9540
tctctcgagc	tttcgcagat	ccgggggggg	aatgagatat	gaaaaagcct	gaactcaccg	9600
cgaagtgagg	cgagaagtgt	ctgatcgaaa	agttcgacag	cgtctccgac	ctgatgcagc	9660
tctcggaggg	cgaagaatct	cgtgctttca	gcttcgatgt	aggaggggcgt	ggatatgtcc	9720
tgcgggtaaa	tagctgcgcc	gatgggttct	acaaagatcg	ttatgtttat	cggcactttg	9780
catcggccgc	gctcccgaat	ccggaaagtc	ttgacattgg	ggagtttagc	gagagcctga	9840
cctattgcat	ctcccgcggt	gcacagggtg	tcacgttgca	agacctgcct	gaaaccgaac	9900
tgcccgctgt	tctacaaccc	gtcgcggagg	ctatggatgc	gatcgtgcg	gccgacttta	9960
gccagacgag	cgggttcggc	ccattcggac	cgcaaggaa	cggccaatac	actacatggc	10020
gtgatttcat	atgcgcgatt	gctgatcccc	atgtgtatca	ctggcaaaact	gtgatggacg	10080
acacccgtcag	tgcgtccgct	gcgcagggtc	tcgatgagct	gatgctttgg	gccgaggact	10140
gcccgaagt	ccggcaccct	gtgcacgcgg	atttcggctc	caacaatgtc	ctgacggaca	10200
atggccgcat	aacagcgggt	attgactgga	gcgaggcgat	gttcggggat	ttcccaatag	10260
aggtcgccaa	catcttcttc	tggaggccgt	gggtggcttg	tatggagcag	cagacgcgct	10320
acttcgagcg	gaggcaatcg	gagcttgacg	gatcgccacg	actccgggcg	tatatgctcc	10380
gcatttggct	tgaccaactc	tatcagagct	tggttgacgg	caatttcgat	gatgcagctt	10440

gggcgaggg	tcatgagcag	gcaatcgctc	gatccggagc	cgggactgtc	gggcgtagac	10500
aaatcgcccc	cagaagcgcg	gccgtctgga	ccgatggctg	tgtagaagta	ctcgccgata	10560
gtggaaacccg	acgccccagc	actcgctccga	gggcaaagaa	atagagtaga	tgccgaccgg	10620
atctgtcgat	cgacaagctc	gagttttctcc	ataataatgt	gtgagtagtt	cccagataag	10680
ggaattagggt	ttcctatagg	gtttcgctca	tgtgttgagc	atataagaaa	cccttagtat	10740
gtattttgtat	ttgtaaaata	cttctatcaa	taaaattttct	aatttcctaaa	acaaaaatcc	10800
agtactaaaa	tccagatccc	ccgaattaat	tccggcgtaa	ttcagatcaa	gcttggcact	10860
ggccgtcgtt	ttacaacgtc	gtgactggga	aaacccctggc	gttaccacca	ttaatcgcc	10920
tgcagcacat	ccccctttcg	ccagctggcg	taatagcgaa	gaggcccgca	ccgatcgccc	10980
ttcccaacag	ttgcgcagcc	tgaatggcga	atgctagagc	agcttgagct	tggtatcgat	11040
tgtcgtttcc	cgccttcagt	ttaaactatc	agtgtttgac	aggatatatt	ggcgggtaaa	11100
cctaagagaa	aagagcggtt	attagaataa	cggatattta	aaagggcggtg	aaaaggttta	11160
tccgttcgtc	catttgtatg	tg				11182

<210> 90
 <211> 8428
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pCambia3300 Plasmid

<400> 90						
catgccaacc	acagggttcc	cctcggggtc	aaagtacttt	gatccaaccc	ctccgctgct	60
atagtgcagt	cggcttctga	cggttcagtgc	agccgtcttc	tgaaaaacgac	atgtcgcaca	120
agtcctaagt	tacgcgacag	gctgccgccc	tgcccttttc	ctggcggttt	cttgtcgctg	180
gttttagtgc	cataaagtag	aatactttgcg	actagaacccg	gagacattac	gccatgaaca	240
agagcgccgc	cgctggcctg	ctggggtatg	ccgcgctcag	caccgacgac	caggacttga	300
ccaaccaacg	ggcgaactg	cacgcggccg	gctgcaccaa	gctgttttcc	gagaagatca	360
ccggcaccag	gcgcgaccgc	ccggagctgg	ccaggatgct	tgaccaccta	cgccctggcg	420
acgtttgtgac	agtgaccagg	ctagaccgcc	tgcccccag	caccgcgac	ctactggaca	480
ttgccgagcg	catccaggag	gccggcgccg	gcctgcgtag	cctggcagag	ccgtgggccc	540
acaccaccac	gccggccggc	cgcatggtgt	tgaccgtgtt	cgccggcatt	gccgagtctg	600
agcggttccct	aatcatcgac	cgcacccgga	gcccggcgga	ggccgccaag	gcccggaggc	660
tgaagtttgg	ccccgcctc	accctcacc	cggcacagat	cgcgcacgce	cgcgagctga	720
tgcaccagga	aggccgcacc	gtgaaagagg	gcctgcaact	gcttggcggtg	catcgctcga	780
ccctgtaccg	cgacttgag	cgcagcgagg	aagtgaacgc	caccgaggcc	aggcggcgcg	840
gtgcctttccg	tgaggacgca	ttgaccggag	ccgacgccct	ggcggccgce	gagaatgaac	900
gccaaagagga	acaagcatga	aaccgcacca	ggacggccag	gacgaaccgt	ttttcattac	960
cgaagagatc	gaggcgga	tgatcgccgc	cgggtacgtg	ttcgagccgc	ccgcgcacgt	1020
ctcaaccgtg	cggctgcatg	aaatcctggc	cggtttgtct	gatgccaagc	tgccggcctg	1080
gccggccagc	ttggcccgctg	aagaaaccca	gcgcgcgctg	ctaaaaaggt	gatgtgtatt	1140
tgagtaaaac	agcttgccgtc	atgcccgtgc	tgcgatatatg	atgcgatgag	taaataaaca	1200
aatacgcaag	gggaacgcat	gaagggtatc	gctgtactta	accagaaagg	cgggtcagge	1260
aagacgacca	tgcgaaccca	tctagcccg	gcctgcaac	tcgcccggggc	cgatgttctg	1320
ttagtcgatt	ccgatcccca	gggcagtgcc	cgcgattggg	cggccgtgcg	ggaagatcaa	1380
ccgctaaccg	ttgtcggcat	cgaccgccc	acgattgacc	gcgacgtgaa	ggccatcggc	1440
cggcgcgact	tcgtagtgat	cgacggagcg	cccagggcg	cggacttggc	tgtgtccgcg	1500
atcaaggcag	ccgacttcgt	gctgattccg	gtgcagccaa	gcccttacga	catatgggcc	1560
accgcccagc	tggtggagct	ggttaagcag	cgcattgagg	tcacggatgg	aaggctacaa	1620
gcggcctttg	tctgttcgctg	ggcgatcaaa	ggcacgcgca	tcggcgggtga	ggttgccgag	1680
gcgctggcgc	ggtacgagct	gcccattctt	gagtcgccga	tcacgcagcg	cgtgagctac	1740
ccaggcactg	ccgcccgcgg	cacaaccgtt	cttgaatcag	aacccgaggg	cgacgctgcc	1800
cgcgaggtcc	aggcgctggc	cgctgaaatt	aaatcaaaac	tcatttgagt	taatgaggtg	1860
aagagaaaat	gagcaaaagc	acaaacacgc	taagtgcggg	ccgtccgagc	gcacgcagca	1920
gcaaggctgc	aacgttggcc	agcctggcag	acacgccagc	catgaagcgg	gtcaactttc	1980
agttgcccgc	ggaggatcac	accaagctga	agatgtacgc	ggtacgcca	ggcaagacca	2040
ttaccgagct	gctatctgaa	tacatcgccg	agctaccaga	gtaaatgagc	aaatgaataa	2100
atgagtagat	gaatttttagc	ggctaaagga	ggcggcatgg	aaaatcaaga	acaaccaggc	2160
accgacgccg	tggaatggcc	catgtgtgga	gggaacggcg	gttggccagg	cgtaagcggc	2220
tgggttgtct	gccggccctg	caatggcact	ggaaccccca	agcccagagg	atcggcgtga	2280
cggtcgcaaa	ccatccggcc	cggtaacaa	cggcgccggc	ctgggtgatg	acctgggtga	2340
gaagttgaag	gccgcgagc	cgcccgagc	gcaacgcac	gaggcagaag	cacgcccggg	2400
tgaatcgtgg	caagcggccg	ctgatcgaat	ccgcaagaa	tcccggcaac	cgccggcagc	2460
cggtgcgccc	tcatattagga	agccgcccga	gggcgacgag	caaccagatt	ttttcggtcc	2520
gatgctctat	gacgtgggca	ccgcgatag	tcgcgacatc	atggacgtgg	ccgttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	gggtgatccg	tacgagcttc	cagacgggca	2640

cgtagagggt	tccgcagggc	cgcccgccat	ggccagtggt	tgggattacg	acctgggtact	2700
gatggcgggt	tooccatctaa	ccgaatccat	gaaccgatac	cgggaaggga	agggagacaa	2760
gcccggccgc	gtgttccgtc	cacacgttgc	ggacgtactc	aagttctgcc	ggcggagccga	2820
tggcggaaaag	cagaaagacg	acctggtaga	aacctgcatt	cgggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cgcccgccctg	gtgacgggtat	ccgaggggtga	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaaaacc	gggcggcccg	agtacatcga	3000
gatcgagcta	gctgattgga	tgtaccgcga	gatcacagaa	ggcaagaacc	cggacgtgct	3060
gacgggtcac	cccgattact	ttttgatcga	tcccggcatc	ggccgttttc	tctaccgcct	3120
ggcaccgccc	gcccagggca	aggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
cagtggcagc	gcccggagagt	tcaagaagtt	ctgtttcacc	gtgcgcaagc	tgatcgggtc	3240
aaatgacctg	ccggagtagc	atttgaagga	caggctgggg	caggctggcc	cgtacctagt	3300
catgcgctac	cgcaacctga	tccagggcga	agcatccgcc	ggttccctaat	gtacggagca	3360
gatgctaggg	caaatgtccc	tagcagggga	aaaaggtcga	aaaggtctct	ttcctgtgga	3420
tagcacgtac	attgggaacc	caaagccgta	cattggggaac	cggaaaccgt	acattgggaa	3480
cccaaagccg	tacattggga	accggtcaca	catgtaagtg	actgatataa	aagagaaaaa	3540
aggcgatttt	tccgcctaaa	actctttaaa	acttattaaa	actcttaaaa	cccgcctggc	3600
ctgtgcataa	ctgctcggcc	agcgcacagc	cgagagcgtg	caaaaagcgc	ctacccttcg	3660
gtcgctgcgc	tccctacgcc	ccgcgcgttc	gcgtcggcct	atcgccggccg	ctggccgcctc	3720
aaaaatggct	ggcctacggc	caggcaatct	accagggcgc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	cgccgcccac	atcaaggcgc	ctgcctcgc	gcgtttcggt	gatgacgggt	3840
aaaacctctg	acacatgcag	ctcccggaga	cggtcacagc	ttgtctgtaa	gcggatgccg	3900
ggagcagaca	agcccgctcag	ggcgcgctcag	cggtgtgttg	cgggtgtcgg	ggcgagcca	3960
tgacccagtc	acgtagcgat	agcggagtg	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgcacc	atatgcgggtg	tgaataaccg	cacagatcgc	taaggagaaa	4080
ataccgcatc	agcgctctct	ccgcttcctc	gctcactgac	tgcgtgcgct	cggctcgttcg	4140
gctgcggcga	gcggctacag	ctcactcaaa	ggcggtaata	cggttatcca	cagaatcagg	4200
ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	4260
ggccgcgttg	ctggcggttt	tccataggct	ccgccccctc	gacgagcatc	acaaaaatcg	4320
acgtcaagt	cagaggtggc	gaaacccgac	aggactataa	agataaccagg	cgtttcccc	4380
tggaaagctc	ctcgtgcgct	ctcctgttcc	gacctgtccg	cttaccggat	acctgtccgc	4440
ctttctccct	tccgggaagcg	tggcgctttc	tcatagctca	cgctgtaggt	atctcagttc	4500
ggtgtaggtc	gttcgctcca	agctgggctg	tgtgcacgaa	cccccgcttc	agccgacctc	4560
ctgcgcctta	tccggttaact	atcgctctga	gtccaacccg	gtaagacacg	acttatcgcc	4620
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	4680
gtcttgaag	tggtggccta	actacggcta	cactagaagg	acagtatttg	gtatctcagg	4740
tctgctgaag	ccagttacct	tccgaaaaag	agttggtagc	tcttgatccg	gcaaacaaac	4800
caccgctggg	agcgggtgggt	tttttggttg	caagcagcag	attacgcgca	gaaaaaaagg	4860
atctcaagaa	gatcctttga	tcttttctac	ggggtctgac	gctcagtga	acgaaaaactc	4920
acgttaaggg	attttgggtca	tgcattctag	gtactaaaaa	aattcatcca	gtaaaaatata	4980
atattttatt	ttctcccaat	caggcttgat	ccccagtaag	tcaaaaaata	gctcgacata	5040
ctgttcttcc	ccgatattct	ccctgatcga	ccggacgcag	aaggcaatgt	cataccatt	5100
gtccgccttc	ccgcttctcc	caagatcaat	aaagccactt	actttggcat	ctttcacaaa	5160
gatgttgctg	tctcccagggt	cgccgtggga	aaagacaagt	tccctcttcg	gcttttccgt	5220
ctttaaaaaa	tcatcacagt	cgccgcgatc	tttaaatgga	gtgtcttctt	cccagtttct	5280
gcaatccaca	tccggccagat	cggtattcag	taagtaatcc	aattcggcta	agcggctgtc	5340
taagctattc	gtatagggac	aatccgatat	gtcgatggag	tgaaagagoc	tgatgcactc	5400
cgcatacagc	tcgataatct	tttcagggct	ttgttcactc	tcatactctt	ccgagcaaa	5460
gacgccatcg	gcccactcca	tgagcagatt	gctccagcca	tcatgccgtt	caaagtgcag	5520
gacctttgga	acaggcagct	ttccttccag	ccatagcatc	atgtcctttt	cccgttccac	5580
atcatagggt	gtccctttat	accggctgtc	cgtcattttt	aaatataggt	tttcattttc	5640
tcccaccagc	ttatatacct	tagcaggaga	cattccttcc	gtatctttta	cgcagcggtg	5700
tttttcgatc	agttttttca	attccgggtga	tattctcatt	ttagccattt	attattttcct	5760
tcctcttttc	tacagtattt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
aattcactgt	tccttgcat	ctaaaacctt	aaataaccaga	aaacagcttt	ttcaaagtgt	5880
ttttcaaagt	tggcgataaa	catagtatcg	acggagccga	ttttgaaacc	gcgggtgatca	5940
caggcagcaa	cgctctgtca	tcgttacaat	caacatgcta	ccctccgcga	gatcatccgt	6000
gtttcaaaac	cggcagctta	gttgccgttc	ttccgaatag	catcggtaac	atgagcaaa	6060
tctgcgcct	tacaacggct	ctcccgctga	cgccgtcccg	gactgatggg	ctgcctgtat	6120
cgagtgggtg	ttttgtgccg	agctgccggg	cggggagctg	ttggctggct	ggtggcagga	6180
tatattgtgg	ttgacgctta	ttgacgctta	gacaacttaa	taacacattg	cggacgtttt	6240
taatgtactg	aattaacgcc	gaattaattc	gggggatctg	gatttttagta	ctggattttg	6300
gttttaggaa	ttagaaattt	tattgataga	agttattttac	aaatacaaat	acatactaag	6360
ggtttcttat	atgctcaaca	catgagcgaa	accctaatg	aacctaatg	cccttatctg	6420
ggaactactc	acacattatt	atggagaaac	tcgagtcaaa	tctcggtgac	gggcaggacc	6480
ggacggggcg	gtaccggcag	gctgaagtcc	agctgccaga	aacccacgtc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgctcgtgc	6600
atgcgcacgc	tccgggtcggt	gggcagcccg	atgacagcga	ccacgctctt	gaagccctgt	6660

```
gcctccaggg acttcagcag gtgggtgtag agcgtggagc ccagtcaccgt ccgctgggtgg 6720
cggggggaga cgtacacggg cgactcggcc gtccagtcgt aggcgttgcg tgccttcacg 6780
gggcccgcgt aggcgatgcc ggcgacctcg ccgtccacct cggcgacgag ccagggatag 6840
cgctcccgcga gacggacgag gtcgtccgtc cactcctgcg gttcctgcgg ctccggtacgg 6900
aagttgaccg tgcttgctc gatgtagtgg ttgacgatgg tgcagaccgc cggcatgtcc 6960
gcctcgggtgg caccggcgat gtcggccggg cgtcgttctg ggctcatggg agactcgaga 7020
gagatagatt tgtagagaga gactgggtgat ttcagcgtgt cctctccaaa tgaaatgaac 7080
aacgatagcc tttcctttat cttgcgaagg atagtgggat tgtgcgtcat cccttacgtc 7140
agtggagata tcacatcaat ccacttgctt tgaagacgtg gttggaacgt cttctttttc 7200
cacgatgtct ctcgtgggtg ggggtccatc tttgggacca ctgtcggcag aggcattctt 7260
aacgatagcc tttcctttat cgcaatgatg gcatttgtag gtgccacctt cctttttctac 7320
tgtccttttg atgaagtgc agatagctgg gcaatggaat ccgaggagggt ttcccgatat 7380
taccctttgt tgaagaagtct caatagccct ttggtcttct gagactgtat ctttgatatt 7440
cttggagtag acgagagtgt cgtgctccac catgttatca catcaatcca cttgctttga 7500
agacgtgggtt ggaacgtctt ctttttccac gatgctcctc gtgggtgggg gtccatcttt 7560
gggaccactg tcggcagagg catcttgaac gatagccttt cctttatcgc aatgatggca 7620
tttgtaggtg ccacttctct tttctactgt ccttttgatg aagtgcagca tagctgggg 7680
atggaatccg agggaggttt ccgatattac cctttgttga aaagtctcaa tagccctttg 7740
gtcttctgag actgtatctt tgatattctt ggagtagacg agagtgtcgt gctccaccat 7800
gttggcaagc tgcttagcc aatacgcmaa cgcctctcc ccgctgttg gccgattcct 7860
taatgcagct ggcacgacag gtttcccgac tggaaagcgg gcagtgcgag caacgcaatt 7920
aatgtgagtt agctcactca ttaggcaccc caggctttac actttatgct tccgggtcgt 7980
atgttggtg agattgtgag cggataacaa tttcacacag gaaacagcta tgaccatgat 8040
tacgaattcg agctcggtag ccgggggatcc tctagagtcg acctgcaggc atgcaagctt 8100
ggcactggcc gtcgttttac aacgtcgtga ctgggaaaac cctggcgtta cccaaactaa 8160
tcgccttgca gcacatccc ctttcgcag agcgaagagg cccgcacga 8220
tcgccttccc caacagttgc gcagcctgaa tggcgaatgc tagagcagct tgagcttggg 8280
tcagattgtc gtttcccgcc ttcagtttaa actatcagtg tttgacagga tatattggcg 8340
ggtaaaccta agagaaaaga gcgtttatta gaataacgga tatttaaaag ggcgtgaaaa 8400
ggtttatccg ttcgtccatt tgtatgtg 8428
```

<210> 91
<211> 3438
<212> DNA
<213> Artificial Sequence

<220>
<223> pLIT38attBZeo Plasmid

```
<400> 91
tcgaccctct agtcaaggcc ttaagtgagt cgtattacgg actggccgctc gttttacaac 60
gctgtgactg ggaiaaccct ggcgttaccc aacttaacgc ccttgacgca catccccctt 120
tcgccagctg gcgtaatagc gaagaggccc gcaccgatcg cccttcccaa cagttgcgca 180
gcctgaatgg cgaatggcgc ttcgcttggt aataaagccc gcttcggcgg gctttttttt 240
gttaactacg tcagggtggc cttttcgggg aaatgtgcgc ggaaccccta tttgtttatt 300
tttctaaata cattcaaata tgtatccgct catgagacaa taaccctgat aaatgcttca 360
ataatattga aaaaggaaga gtatgagtat tcaacatttc cgtgtcgccc ttattccctt 420
ttttgcggca ttttgccttc ctgtttttgc tcacccagaa acgctgggtg aagtaaaaga 480
tgctgaagat cagttgggtg caccagtggtg ttacatcgaa ctggatctca acagcggtaa 540
gatccttgag agttttcgcc ccgaagaacg ttctccaatg atgagcactt ttaaagttct 600
gctatgtggc gcggtattat cccgtgttga cgcggggcaa gagcaactcg gtcgccgat 660
acactattct cagaatgact tgggttgagta ctcaccagtc acagaaaagc atcttacgga 720
tggcatgaca gtaagagaat tatgcagtgc tgccataacc atgagtgata aactgcggc 780
caacttactt ctgacaacga tcggaggacc gaaggagcta accgcttttt tgcacaacat 840
gggggatcat gtaactcgcc ttgatcggtg ggaaccggag ctgaatgaag ccataccaaa 900
cgacgagcgt gacaccacga tgctgtagc aatggcaaca acgttgcgca aactattaac 960
tggcgaacta cttactctag cttcccggca acaattaata gactggatgg aggcggataa 1020
agttgcagga ccacttctgc gctcggccct tccggctggc tgggtttattg ctgataaatc 1080
tggagccggt gagegtgggt ctcgcggtat catgacgca ctggggccag atggtaagcc 1140
ctcccgatc tcagttatct acacgacggg gactcaggca actatggatg aacgaaatag 1200
acagatcgct gagataggtg cctcactgat taagcattgg taactgtcag accaagttta 1260
ctcatatata ctttagattg atttaccctg gttgataatc agaaaagccc caaaaacagg 1320
aagattgtat aagcaaatat ttaaattgta aacgttaata tttgttaaa attcgcgtta 1380
aatttttggt aaatcagctc attttttaac caataggccg aaatcggcaa aatcccttat 1440
aaatcaaaag aatagccgca gatagggttg agtgttgttc cagtttggaa caagagcca 1500
ctattaaaga acgtggactc caacgtcaaa cggctctatc cgcgtgatgg 1560
ccactacgtg aaccatcacc caaatcaagt tttttggggg cgagggtgccg taaagcacta 1620
```

```
aatcggaacc ctaaagggag cccccgattt agagcttgac ggggaaagcg aacgtggcga 1680
gaaaggaagg gaagaaaagc aaaggagcgg gcgctagggc gctggcaagt gtagcgggtca 1740
cgctgcgcgt aaccaccaca cccgcgcgcg ttaatgcgcc gctacagggc gcgtaaaagg 1800
atctaggtga agatcctttt tgataatctc atgaccaaaa tcccttaacg tgagtttttcg 1860
ttccactgag cgtcagaccc cgtagaaaag atcaaaaggat cttcttgaga tccttttttt 1920
ctgcgcgtaa tctgctgctt gcaaacaaaa aaaccaccgc taccagcggg ggtttgtttg 1980
ccggatcaag agctaccaac tctttttccg aaggtaactg gcttcagcag agcgcagata 2040
ccaaatactg ttcttctagt gttagccctag ttaggccacc acttcaagaa ctctgtagca 2100
ccgcctacat acctcgctct gctaatacctg ttaccagtgg ctgctgccag tggcgataag 2160
tcgtgtctta cgggtttgga ctcaagacga tagttaccgg ataaggcgca gcggtcgggc 2220
tgaacggggg gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga 2280
tacctacagc gtgagctatg agaaagcgcc acgcttcccg aagggagaaa ggcggacagg 2340
tatccggtaa gcggcagggg cggaacagga gagcgcacga gggagcttcc agggggaaac 2400
gcctggtatc tttatagttc tgtcgggttt cgccacctct gacttgagcg tcgatttttg 2460
tgatgctcgt cagggggggc gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg 2520
ttcctggcct tttgctggcc ttttgctcac atgtaatgtg agttagctca ctcattaggc 2580
acccaggctt ttacacttta tgcttccggc tcgtatgttg tgtggaattg tgagcgata 2640
acaattttcac acaggaaaca gctatgacca tgattacgcc aagctacgta atacgactca 2700
ctagtggggc ccgtgcaatt gaagccggct ggcgccaagg ttctctgcag gattgaagcc 2760
tgctttttta tactaacttg agcgaaatct ggcctcatgg ccaagttgac cagtgcctt 2820
ccggtgctca ccgcgcgcga cgtcgcggga gcggtcgagt tctggaccga ccggctcggg 2880
ttctcccggg acttcgtgga ggacgacttc gccggtgtgg tccgggacga cgtgacctg 2940
ttcatcagcg cggtccagga ccaggtggtg cggacaaca tccgggctg ggtgtgggtg 3000
cgcgccctgg acgagctgta cgccgagttg tccgaggtcg ggtgtgggtg cttccgggac 3060
gcctccgggc cgcccatgac cgagatcggc gagcagccgt gggggcgggg gttcgccctg 3120
cgcgaccccg ccggcaactg cgtgcacttc gtggccgagg agcaggactg acacgtgcta 3180
cgagatttcg attccaccgc gccttctat acagcgcggg cgttttcggg 3240
gacgcgggct ggatgatcct ccagcgcggg gatctcatgc tggagttctt cgcccacccc 3300
aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca 3360
aataaagcat ttttttact gcattctagt tgtggtttgt ccaaaactcat caatgtatct 3420
tatcatgtct gtataccc 3438
```

<210> 92
<211> 10549
<212> DNA
<213> Artificial Sequence

<220>
<223> pCambia1302 Plasmid

<300>
<308> Genbank #AF234398
<309> 2000-04-24

```
<400> 92
catggtagat ctgactagta aaggagaaga acttttctact ggagttgtcc caattcttgt 60
tgaattagat ggtgatgtta atgggcacaa attttctgtc agtggagagg gtgaagggtga 120
tgcaacatac ggaaaactta cccttaaatt tatttgcact actggaaaac tacctgttcc 180
gtggccaaca cttgtcacta ctttctctta tgggtgttcaa tgcttttcaa gatacccca 240
tcatatgaag cggcacgact tcttcaagag cgccatgcct gagggatacg tgcaggagag 300
gaccatcttc ttcaaggacg acgggaacta caagacacgt gctgaagtca agtttgaggg 360
agacaccctc gtcaacagga tctgagcttaa ggaatcgat ttcaaggagg acggaaacat 420
cctcgccac aagttggaat acaactacaa ctcccacaac gtatacatca tggccgacaa 480
gcaaaagaac ggcataaaag ccaacttcaa gaccgccac aacatcgaag acggcgcgct 540
gcaactcgct gatcattatc aacaaaatc tccaattggc gatggccctg tccttttacc 600
agacaaccat tacctgtcca cacaatctgc cctttcgaaa gatcccaacg aaaagagaga 660
ccacatggtc cttcttgagt ttgtaacagc tgctgggatt acacatggca tggatgaact 720
atacaaagct agccaccacc accaccacca cgtgtgaatt ggtgaccagc tcgaatttcc 780
ccgatcgttc aaacatttgg caataaagtt tcttaagatt gaatcctgtt gccggtcttg 840
cgatgattat ctgttgaatt acgttaagca tgtaataatt aacatgtaac tggatgaact 900
gcatgacgtt atttatgaga tgggttttta tgattagagt cccgcaatta tacatttaat 960
acgcgataga aaacaaaata tagcgcgcaa attatgcgcg attatcgcgc gcggtgtcat 1020
ctatgttact agatcgggaa ttaaaactat agtggttgac aggatataat ggcgggtaaa 1080
cctaagagaa aagagcgttt attagaataa cggatattta aaagggcggtg aaaaggttta 1140
tccgttcgtc catttgtatg tgcatgccaa ccacagggtt cccctcgga tcaaagttact 1200
ttgatccaac ccttcgctg ctatagtgca gtcggcttct gacgttcagt gcagcgtct 1260
tctgaaaacg acatgtcgca caagtcctaa gttacgcgac aggctgccgc cctgcccttt 1320
```

tcctggcggtt	ttcttgtcgc	gtgttttagt	cgcataaagt	agaataacttg	cgactagaac	1380
cggagacatt	acgccatgaa	caagagcgcc	gccgctggcc	tgctgggcta	tgcccgcgtc	1440
agcaccgacg	accaggactt	gaccaaccaa	cggggccgaac	tgcaocgccc	cggctgcacc	1500
aagctgtttt	ccgagaagat	caccggcacc	aggcgcgacc	gcccggagct	ggccaggatg	1560
cttgaccacc	tacgccctgg	cgacgttgtg	acagtgaacca	ggctagaccg	cctggcccgc	1620
agcaccgcg	acctactgga	cattgccgag	cgcatccagg	aggccggcgc	gggctgcgt	1680
agcctggcag	agccgtgggc	cgacaccacc	acgcccggccg	gcccgcattg	gttgaccgtg	1740
ttcgccggca	ttgcccaggt	cgagcgttcc	ctaatacatcg	accgcacccg	gagcgggccc	1800
gaggccgcat	aggcccaggg	cgtgaagttt	ggcccccgcc	ctacccctcac	cccggcacag	1860
atcgccgacg	cccgcgagct	gatcgaccag	gaaggccgca	ccgtgaaaga	ggcggctgca	1920
ctgcttggcg	tgcatcgctc	gaccctgtac	cgcgcacttg	agcgcagcga	ggaagtgcg	1980
cccaccggag	ccaggcggcg	cgggtgcctt	cgtgaggacg	cattgaccga	ggccgacgcc	2040
ctggcggccg	ccgagaatga	acgccaagag	gaacaagcat	gaaaccgcac	caggacggcc	2100
aggacgaacc	gtttttcatt	accgaagaga	tgcaggcgga	gatgatcgcg	gccgggttac	2160
tgctcgagcc	gcccgcgcac	gtctcaaccg	tgccgctgca	tgaaatcctg	gccggtttgt	2220
ctgatgccaa	gctggcggcc	tgcccgccca	gcttggccgc	tgaagaaacc	gagcggccgc	2280
gtctaaaaag	gtgatgtgta	tttgagtaaa	acagcttgcg	tcatgcggtc	gctgcgtata	2340
tgatgcatg	agtaataaaa	caaatacga	aggggaacgc	atgaaggtta	tcgctgtact	2400
taaccagaaa	ggcgggtcag	gcaagacgac	catcgcaacc	catctagccc	gcccctgca	2460
actcgccggg	gcccgatgtt	tgtagtoga	ttccgatccc	cagggcagtg	cccgcgattg	2520
ggcggccggt	ccgggaagat	aaccgctaac	cgttgctcgg	atcgaccgcc	cgacgattga	2580
ccgcgacgtg	aaggccatcg	gccggcgcga	cttcgtagtg	atcgaccggg	cgccccaggc	2640
ggcggacttg	gctgtgtccg	cgatcaaggc	agccgacttc	gtgctgattc	cggtgcagcc	2700
aagcccttac	gacatattgg	ccaccgcccga	cctgggtggg	ctgggttaagc	agcgcattga	2760
ggtcacggat	ggaaggctac	aagcggcctt	tgctgtgtcg	cgggcgatca	aaggcacgcg	2820
catcgccggt	gaggttgccg	aggcgctggc	cgggtacgag	ctgcccattc	ttgagtcgcc	2880
tatcacgcag	cgctgagct	accagggcac	tgccgcccgc	ggcacaaccg	ttcttgaatc	2940
agaaccggag	ggcgacgctg	cccgcgaggt	ccaggcgctg	gccgctgaaa	ttaaatcaaa	3000
actcatttga	gttaatgagg	taaaagaaaa	atgagcaaaa	gcacaaacac	gctaagtgcc	3060
ggccgtccga	ccgcacgcag	cagcaaggct	gcaacgcttg	ccagcctggc	agacacgcca	3120
gcatgaagc	gggtcaactt	tcagttgccg	gcggaggatc	acaccaagct	gaagatgtac	3180
gcggtacgcc	aaggcaagac	cattaccgag	ctgctatctg	aatacatcgc	gcagctacca	3240
gagtaaatga	gcaaatgaat	aaatgagtag	atgaatttta	gcccgtaaaag	gaggcgcat	3300
ggaaaaatcaa	gaacaaccag	gcaccgacgc	cgtggaatgc	cccatgtgtg	gaggaaacggg	3360
cgggtggcca	ggcgtaagcg	gctgggttgt	ctgcggcccc	tgcaatggca	ctggaaacccc	3420
caagcccgca	gaactgcgtg	gacggtcgcc	aaccatccgg	cccggtaaaa	atccgcccgg	3480
cgctgggtga	tgacctgggt	gagaagttga	aggccgcgca	ggccgcccag	cggcaacgca	3540
tcgaggcaga	agcacgcccc	ggtgaatcgt	ggcgaagcgc	cgctgatcga	atccgcaaaag	3600
aatcccgcca	ttttttcggt	cccggtcgcc	cgtcgattag	gaagccgccc	aaggcgagc	3660
agcaaccaga	ttttttcggt	ccgatgctct	atgacgtggg	caccgcgcat	agtgcgagca	3720
tcattggacgt	ggcggttttc	cgtctgtcga	agcgtgaccg	acgagctggc	gaggtgatcc	3780
gctacgagct	tcacagcggg	cacgtagagg	tttcgcagg	gcccggccggc	atggccgagc	3840
gtggggatta	cgacctgggt	ctgatggcgg	tttcccatct	aaccgaatcc	atgaaccgat	3900
accgggaagg	gaaggagagc	aagcccggcc	gcgtgttccg	tcacacagct	gcggacgtac	3960
tcaagttctg	ccggcgagcc	gatggcggaa	agcagaaga	cgacctggtg	gaaacatgca	4020
ttcggttaaaa	caccacgcac	gttgccatga	agcgtacgaa	gaaggccaag	aacggccgcc	4080
tggtgacggt	atccgagggc	gaagccttga	ttagccgcta	caagatcgta	aagagcgaaa	4140
ccgggcccgc	ggagtacatc	gagatcgagc	tagctgattg	gatgtaccgc	gagatcacag	4200
aaggcaagaa	cccggacgtg	ctgacggttc	accccgatta	ctttttgatc	gatcccggca	4260
tcggccggtt	tctctaccgc	ctggcacgcc	gcgcccaggg	caaggcagaa	gccagatggg	4320
tggtcaagac	gatctacgaa	cgagtgggca	gcgcccggga	ggtcaagaag	ttctgtttca	4380
ccgtgcgcaa	gctgatcggg	tcaaatgacc	tgccggagta	cgatttgaag	gaggaggcgg	4440
ggcaggctgg	cccgatccca	gtcatgcgct	accgcaacct	gatcgagggc	gaagcatccg	4500
ccggttccta	atgtacggag	cagatgctag	ggcaaatggc	cctagcaggg	gaaaaaggte	4560
gaaaaggctt	ctttcctgtg	gatagcacgt	acattgggaa	cccaaagccg	tacattggga	4620
accggaaccc	gtacattggg	aacccaaaagc	cgtacattgg	gaaccgggtc	cacatgtatg	4680
tgactgatat	aaaagagaaa	aaaggcgatt	tttccgccta	aaactcttta	aaacttatta	4740
aaactcttaa	aaccgcctg	gcctgtgcct	aactgtctgg	ccagcgcaca	gccgaagagc	4800
tgcaaaaagc	gcctaccctt	cggtcgctgc	gctccctacg	ccccgcccgt	tcgctgcggc	4860
ctatcgccgc	cgctggccgc	tcaaaaatgg	ctggcctacg	gccaggcaat	ctaccaggcc	4920
gcggagcagc	cgcgcgcgtg	ccactcgacc	ccggcgcccc	acatcaaggg	acccctcctc	4980
gcgcggtttc	gtgatgacgg	tgaaaaacct	tgacacatgc	agctcccggg	gacgggtcaca	5040
gcttgtctgt	aagcggatgc	cgggagcaga	caagcccgct	agggcgcgct	agcgggtgtt	5100
ggcgggtgtc	ggggccgagc	catgaccagc	tcacgttagc	atagcggagt	gtatagttgg	5160
ttactatgc	ggcatcagag	cagattgtac	tgagagtgc	ccatatgcgg	tgtgaaatag	5220
cgcacagatg	cgttaaggaga	aaataccgca	tcaggcgctc	ttccgcttcc	tcgctcactg	5280
actcgctgcg	ctcggtcggt	cggctgcggc	gagcgggtac	agctcactca	aaggcggtaa	5340

tacgggttatc	cacagaatca	ggggataacg	caggaaagaa	catgtgagca	aaaggccagc	5400
aaaaggccag	gaaccgtaaa	aaggccgcgt	tgctggcggt	tttccatagg	ctccgcccc	5460
ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagagggtg	gcgaaacccg	acaggactat	5520
aaagatacca	ggcggtttccc	cctggaagct	ccctcggtgcg	ctctcctggt	ccgaccctgc	5580
cgcttaccgg	atacctgtcc	gcctttctcc	cttcggggaag	cgtggcgctt	tctcatagct	5640
cacgctgtag	gtatctcagt	tcgggtgtagg	tcgttcgctc	caagctgggc	tgtgtgcacg	5700
aaccccccg	tcagccccgac	cgctgcgcct	tatccggtaa	ctatcgtctt	gagtccaacc	5760
cggtaaagaca	cgacttatcg	ccactggcag	cagccactgg	taacaggatt	agcagagcga	5820
gggtatgtagg	cggtgctaca	gagttcttga	agtgggtggcc	taactacggc	tacactagaa	5880
ggacagtatt	tgggtatctgc	gctctgctga	agccaggttac	cttcggaaaa	agagttggta	5940
gctcttgatc	cggcaaacaa	accaccgctg	gtagcggtagg	tttttttgg	tgcaagcagc	6000
agattacgcg	cagaaaaaaa	ggatctcaag	aagatccctt	gatcttttct	acgggggtctg	6060
acgctcagtg	gaacgaaaaa	tcacgttaag	ggatttttgg	catgcattct	aggtactaaa	6120
acaattcatc	cagtaaaaaa	taatatttta	ttttctccca	atcaggcttg	atccccagta	6180
agtcaaaaaa	tagctcgaca	tactgttctt	ccccgatata	ctccctgatc	gaccggacgc	6240
agaaggcaat	gtcataccac	ttgtccgccc	tgccgcttct	cccaagatca	ataaagccac	6300
ttactttgce	atctttcaca	aagatgttgc	tgctctccag	gtcgcggtgg	gaaaaagaca	6360
gttcctcttc	gggcttttaa	gtcttttaaa	aatcatacag	ctcgcgcgga	tcttttaaatg	6420
gagtgctctc	ttcccagttt	tcgcaatcca	catcggccag	atcggtattc	agtaagtaat	6480
ccaattcggc	taagcggctg	tctaagctat	tcgtataagg	acaatccgat	atgtcagtagg	6540
agtgaagag	cctgatgcac	tcgcataaca	gctcgataat	cttttcaggg	ctttgttcat	6600
cttcatactc	ttccgagcaa	aggacgccat	cggcctcact	catgagcaga	ttgctccagc	6660
catcatgccg	ttcaaagtgc	aggacctttg	gaacaggcag	ctttccttcc	agccatagca	6720
tcagtgtctt	ttcccgttcc	acatcatagg	tggtcccttt	ataccggctg	tccgtcattt	6780
ttaaatatag	gttttcatatt	tctcccacca	gcttatatac	cttagcagga	gacattcctt	6840
ccgtatcttt	tacgcagcgg	tatttttctg	tcagtttttt	caattccggg	gatatttctca	6900
tttagctcat	ttattatttt	cttcctcttt	ttcacagtat	ttaaagatac	cccaagaagc	6960
taattataac	aagacgaact	ccaattcact	gttccctgca	ttctaaaaac	ttaaatacca	7020
gaaaacagct	ttttcaaagt	tgttttcaaa	gttggcggtat	aacatagtat	cgacggagcc	7080
gattttgaaa	cccggttgat	cacaggcagc	aacgctctgt	catcggtaca	atcaacatgc	7140
taccctccgc	gagatcatcc	gtgtttcaaa	cccggcagct	tagttgcccg	tcttccgaat	7200
agcatcggtg	acatgagcaa	agtctgccgc	cttacaacgg	ctctcccgtc	gacgccgtcc	7260
cggactgagt	ggctgcctgt	atcgagtggt	gattttgtgc	cgagctgccg	gtcgggggag	7320
tggtggctgg	ctgggtggcag	gatataattgt	ggtgtaaaca	aattgacgct	tagacaactt	7380
aataacacat	tgccggacgtt	tttaatgtac	tgaattaacg	ccgaattaat	tcggggggatc	7440
tggattttag	tactggattt	tggttttagg	aattagaaat	tttattgata	gaagtatttt	7500
acaaatacaa	atacatacta	agggtttctt	atatgtccta	cacatgagcg	aaaccctata	7560
ggaaccctaa	ttcccttatc	tgggaactac	tcacacatta	ttatggagaa	actcgagctt	7620
gtcagatcag	gtcgatcggt	ggcatctact	ctatttcttt	gccctcggac	gagtggtggg	7680
gcgtcggttt	ccactatcgg	cgagtacttc	tacacagcca	tcggtccaga	cggccgcgct	7740
tctgcgggcg	atttgtgtac	gcccgcagat	cccggctccg	gatcggacga	ttgcgtcgca	7800
tcgacctgac	gcccgaagctg	catcatcgaa	attgcgctca	accaagctct	gatagagttg	7860
gtcaagacca	atgcggagca	tatacgccc	gagtcgtggc	gatcctgcaa	gctccggatg	7920
cctccgctcg	aagtagcgcg	tctgctgctc	catacaagcc	aaccacggcc	tccagaagaa	7980
gatgtttggc	acctcgattt	gggaatcccc	gaacatcgcc	tcgctccagt	caatgaccgc	8040
tggtatgccc	ccattgtccg	tcaggacatt	gttgagccg	aaatccgcgt	gcacgaggtg	8100
cgggacttcg	gggcagtcct	cggcccaaa	catcagctca	tcgagagcct	gcgcgacgga	8160
cgactgacg	gtgtcgcca	tcacagtttg	ccagtgtatac	acatggggat	cagcaatcgc	8220
gcatatgaaa	tcacgccatg	tagtgtattg	accgattcct	tgcggtccga	atgggcccga	8280
cccgtctgtc	tgggtaaagt	cggccgcagc	gatcgcattc	atagcctccg	cgaccgggtg	8340
tagaacagcg	ggcagttcgg	tttcaggcag	gtcttgcaac	gtgacaccct	gtgcacggcg	8400
ggagatgcaa	taggtcaggg	tctcgctaaa	ctccccaatg	tcaagcactt	ccggaatcgg	8460
gagcgcggcc	gatgcaaaat	gcccataaac	ataacgatct	ttgtagaaac	catcggcgca	8520
gctattttacc	cgcaggacat	atcccagccc	tcctacatcg	aagctgaaa	cacgagattc	8580
ttcgccctcc	gagagctgca	tcaggctcga	gacgctgtcg	aacttttctg	tcagaaactt	8640
ctcgacagac	gtcgcgggtg	gttcaggctt	tttcatatct	cattgcccc	cgggatctgc	8700
gaaagactga	gagagataga	ttttagagag	gagactgggtg	atttcagcgt	gtcctctcca	8760
aatgaaatga	acttccttat	atagaggaag	gtcttgcgaa	ggatagtggg	attgtgcgtc	8820
atcccttacg	tcagttggaga	tatcacatca	atccacttgc	tttgaagacg	tggttggaac	8880
gtcttctttt	ttccacgatgc	tctcgtggg	tgggggtcca	tctttggggac	cactgtcgcc	8940
agaggcatct	tgaacgatag	ccttttcttt	atcgcaatga	tggcatttgt	aggtgccacc	9000
ttccttttct	actgtccttt	tgatgaaagt	acagatagct	gggcaaatgga	atccgaggag	9060
gtttcccgat	attacccttt	gttgaaaagt	ctcaatagcc	ctttgggtctt	ctgagactgt	9120
attctttgata	ttctttggagt	agacgagagt	gtcgtgctcc	accatgttat	cacatcaatc	9180
cacttgcttt	gaagacgtgg	ttggaaacgtc	ttctttttcc	acgatgctcc	tcgtgggttg	9240
gggtccatct	ttgggaccac	tgtcggcaga	ggcatctga	acgatagcct	ttcctttatc	9300
gcaatgatgg	cattttgtagg	tggccacttc	cttttctact	gtccttttga	tgaagtgaca	9360

gatagctggg	caatggaatc	cgaggaggtt	tccgatatt	accctttgtt	gaaaagtctc	9420
aatagccctt	tgggtcttctg	agactgtatc	tttgatattc	ttggagtaga	cgagagtgtc	9480
gtgctccacc	atgtttggcaa	gctgctctag	ccaatacgc	aaccgcctct	ccccgcgcgt	9540
tggccgattc	attaatgcag	ctggcacgac	agggtttccg	actggaaagc	gggcagttag	9600
cgcaacgcaa	ttaatgtgag	ttagctcact	cattaggcac	cccaggcttt	acactttatg	9660
cttccggctc	gtatgtttgtg	tgggaattgtg	agcggataac	aatttcacac	aggaaacagc	9720
tatgaccatg	attacgaatt	cgagctcggg	acccggggat	cctctagagt	cgacctgcag	9780
gcatgcaagc	ttggcactgg	ccgtcgtttt	acaacgtcgt	gactgggaaa	accctggcgt	9840
taccgaactt	aatcgccttg	cagcacatcc	ccctttcgcc	agctggcgta	atagcgaaga	9900
ggcccgacc	gatcgccctt	cccaacagtt	gcgagcctg	aatggcgaat	gctagagcag	9960
cttgagcttg	gatcagattg	tcgtttcccg	cccttcagttt	agcttcattg	agtcaaagat	10020
tcaaatagag	gacctaacag	aactcgccgt	aaagactggc	gaacagtcca	tacagagtct	10080
cttacgactc	aatgacaaga	agaaaatctt	cgtcaacatg	gtggagcacg	acacacttgt	10140
ctactccaaa	aatatcaaa	atacagtcctc	agaagaccaa	agggcaattg	agacttttca	10200
acaaagggtg	atatccggaa	acctcctcgg	attccattgc	ccagctatct	gtcactttat	10260
tgtgaagata	gtggaaaagg	aaggtggctc	ctacaaatgc	catcattgag	ataaaggaaa	10320
ggccatcggt	gaagatgcct	ctgccgacag	tggtcccaaa	gatggacccc	caccacagag	10380
gagcatcgtg	gaaaaagaag	acgttccaac	cacgtcttca	aagcaagtgg	attgatgtga	10440
tatctccact	gacgtaagg	atgacgcaca	atcccaactat	ccttcgcaag	acccttcctc	10500
tatataagga	agttcatttc	atttgagag	aacacggggg	actcttgac		10549

<210> 93
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> CaMV35SpolyA Primer

<400> 93
ctgaattaac gccgaattaa ttcgggggat ctg 33

<210> 94
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> CaMV35Spr Primer

<400> 94
ctagagcagc ttgccaacat ggtggagca 29

<210> 95
<211> 12592
<212> DNA
<213> Artificial Sequence

<220>
<223> pAg2 Plasmid

<400> 95	gtacgaagaa	ggccaagaac	ggccgcctgg	tgacgggtatc	cgagggtgaa	gccttgatta	60
	gccgctacaa	gatcgtaaa	agcgaaaaccg	ggcggccgga	gtacatcgag	atcgagctag	120
	ctgattggat	gtaccgagag	atcacagaag	gcaagaaccc	ggacgtgctg	acggttcacc	180
	ccgattactt	tttgatcgat	cccggcacgc	gccgttttct	ctaccgcctg	gcacgccgcg	240
	ccgcaggcaa	ggcagaagcc	agatgggtgt	tcaagacgat	ctacgaacgc	agtggcagcg	300
	ccggagagtt	caagaagttc	tgtttcaccc	tgcgcaagct	gatcgggtca	aatgacctgc	360
	cggagtacga	tttgaaggag	gaggcggggc	aggctggccc	gacccatgct	atgcgctacc	420
	gcaacctgat	cgaggcgcaa	gcacccgcgc	gttcctaagt	tacggagcag	atgctagggc	480
	aaattgccct	agcaggggaa	aaaggtcgaa	aagggtctct	tcctgtggat	agcacgtaca	540
	ttgggaaccc	aaagccgtac	attgggaacc	ggaacccgta	catttggaac	ccaaagccgt	600
	acattgggaa	ccggtcacac	atgtaagtga	ctgatataaa	agagaaaaaa	ggcgattttt	660
	ccgcctaaaa	ctcttttaaa	cttattaaaa	ctcttaaaac	ccgcctggcc	tgtgcataac	720
	tgtctggcca	gcgcacagcc	gaagagctgc	aaaaagcgcc	tacccttcgg	tcgctgcgct	780
	ccctacggcc	cgccgcttcg	cgctggccta	tcggcgccgc	tgcccgctca	aaaatggctg	840
	gcctacggcc	aggcaatcta	ccaggcgccg	gacaagccgc	gccgtcgcca	ctcgaccgcc	900

ggcgccaca	tcaaggcacc	ctgcctcgcg	cgtttcggtg	atgacgggtga	aaacctctga	960
cacatgcagc	tcccggagac	gggtcacagct	tgtctgtaag	cggtatgccgg	gagcagacaa	1020
gcccgtcagg	gcgcgtcagc	gggtgttggc	gggtgtcggg	gcgcagccat	gacccagtc	1080
cgtagcgata	gcggagtgtg	tactggctta	actatgccc	atcagagcag	attgtactga	1140
gagtgcacca	tatgcgggtg	gaaataccgc	acagatgcgt	aaggagaaaa	taccgcatca	1200
ggcgctcttc	cgcttctctg	ctcactgact	cgctgcgctc	ggctcgttcgg	ctgcggcgag	1260
cggtatcagc	tcactcaaag	gcggtaatat	ggttatccac	agaatcaggg	gataacgcag	1320
gaaagaacat	gtgagcaaaa	ggccagcaaa	aggccaggaa	ccgtaaaaag	gccgcgttgc	1380
tggcggtttt	ccataggctc	cgccccctg	acgagcatca	caaaaaatcg	cgctcaagtc	1440
agaggtggcg	aaaccgcaca	ggactataaa	gataaccaggc	gtttccccct	ggaagctccc	1500
tcgtgcgctc	tcctgttccg	accctgcgcg	ttaccggata	cctgtccgcc	ttctctccct	1560
cggaagcggt	ggcgctttct	catagctcac	gctgtaggta	tctcagttcg	gtgtagggtcg	1620
ttcgctccaa	gctgggctgt	gtgcacgaac	cccccgttca	gcccagccgc	tgcgccttat	1680
ccggttaacta	tcgtcttgag	tccaaccggg	taagacacga	cttatcgcca	ctggcagcag	1740
ccactggtaa	caggattagc	agagcgagggt	atgtaggcgg	tgctacagag	ttcttgaagt	1800
gggtggcctaa	ctacggctac	actagaagga	cagtatttgg	tatctgcgct	ctgctgaagc	1860
cagttacctt	cggaaaaaga	gttggtagct	cttgatccgg	caaacaaaacc	accgctggta	1920
gcggtgggtt	ttttgtttgc	aagcagcaga	ttacgcgcag	aaaaaaagga	tctcaagaag	1980
atcctttgat	cttttctacg	gggtctgacg	ctcagtgga	cgaaaactca	cgtaagggga	2040
ttttgctcat	tccttctagg	tactaaaaca	ttaccggata	taaaatataa	tattttatct	2100
tctcccaatc	aggcttgatc	cccagtaagt	caaaaaatag	ctcgacatac	tgttcttccc	2160
cgatatcctc	cctgatcgac	cggacgcaga	aggcaatgtc	ataccacttg	tccgccttgc	2220
cgcttctccc	aagatcaata	aagccactta	ctttgccatc	tttcacaaag	atgttgctgt	2280
ctcccaggtc	gccgtgggaa	aagacaagtt	cctcttcggg	cttttccgct	tttaaaaaat	2340
catacagctc	gcgcggatct	ttaaatggag	tgtcttcttc	ccagttttcg	caatccacat	2400
cggccagatc	gttattcagt	aagtaatcca	attcggctaa	gcggctgtct	aagcttctcg	2460
tatagggaca	atccgatatg	tcgatggagt	gaaagagcct	gatgcactcc	gcatacagct	2520
cgataatctt	ttcagggtct	tgttcatctt	catactcttc	cgagcaaaagg	acgccatcgg	2580
cctcatcatc	gttcaggatg	ctccagccat	catgcctgtc	aaagtgcagg	acctttggaa	2640
caggcagctt	tccttccagc	catagcatca	tgtccttttc	ccgttccaca	tcatagggtg	2700
tccttttata	ccggtgttcc	gtcattttta	aatataggtt	ttcattttct	cccaccagct	2760
tatatccact	agcaggagac	attccttccg	ctttgtttac	gcagcggtat	ttttcgctca	2820
gttttttcaa	ttccggtgat	attctcatct	tagccattta	ttatttccct	cctcttttct	2880
acagatttta	aagatacccc	aagaagctaa	ttataacaag	acgaactcca	attcactgtt	2940
ccttgcatct	taaaacctta	aataaccagaa	aacagctttt	tcaaagttgt	tttcaaagct	3000
ggcgtataaac	atagtatcga	cggagccgat	tttgaaaccg	cggtgatcac	aggcagcaac	3060
gctctgtcat	cgttacaatc	aacatgctac	cctccgcgag	atcatccgtg	tttcaaaacc	3120
ggcagcttag	ttgcgcttct	tccgaatagc	actcggtaaa	tgagcaaaagt	ctgcgcctct	3180
acaacggctc	ttccgctgac	gccgtcccgg	actgatgggc	tgccctgtatc	gagtgtgtat	3240
tttgtgcgga	gctgcgggtc	ggggagctgt	tggtcggtcg	gtggcaggat	atattgtggt	3300
gtaaacaaat	tgacgcttag	acaacttaat	aacacattgc	ggacgttttt	aatgtactga	3360
attaacgccg	aattaattcg	ggggatctgg	attttagtac	tggattttgg	tttttaggaat	3420
tagaaatttt	attgatagaa	gtattttaca	aatacaaaat	cataactaagg	gtttcttata	3480
tgctcaacac	atgagcgaaa	ccctatagga	accctaattc	ccttatctgg	gaactactca	3540
cacattatta	tgagaaaact	cgagtcacat	ctcggtgacg	ggcaggagccg	gacggggcgg	3600
taccggcagg	ctgaagtcca	gctgccagaa	acccacgtca	tgccagttcc	cgtgcttgaa	3660
gccggccgcc	cgcagcatgc	cgcggggggc	atatccgagc	gcctcgtgca	tgcgcacgct	3720
cggtcggttg	ggcagcccg	tgacagcgac	cacgctcttg	aagccctgtg	cctccaggga	3780
cttcagcagg	tgggtgtaga	gcgtggagcc	cagtcctcgt	cgctgggtggc	ggggggagac	3840
gtacacggct	gactcggccg	tccagtcgta	ggcgttgctg	gccttccagg	ggcccgcgta	3900
ggcgatgccc	gcgacctcgc	cgtccacctc	ggcgacgagc	cagggatagc	gctcccgag	3960
acggacgagg	tcgtccgctc	actcctgcgg	ttcctgcggc	tcggtacgga	agttgacgct	4020
gcttgtctcg	atgtagtggg	tgacgatggg	gcagaccgcc	ggcatgtccg	cctcggtggc	4080
acggcggtatg	tcggccgggc	gtcggtctgg	gctcatggta	gactcgagag	agatagattt	4140
gtagagagag	actggtgatt	tcagcgtgtc	ctctccaaat	gaaatgaact	tccttatata	4200
gaggaagggt	ttgcgaagga	tagtgggatt	gtgcgtcatc	ccttacgtca	gtggagatat	4260
cacatcaatc	cacttgcttt	gaagacgtgg	ttggaaacgtc	ttctttttcc	acgatgctcc	4320
tcgtgggtgg	gggtccatct	ttgggaccac	tgtcggcaga	ggcatcttga	acgatagcct	4380
ttcctttatc	gcaatgatgg	cattttagtg	tgccaccttc	cttttctact	gtccttttga	4440
tgaagtgcaca	caatggatgg	cgaggaggtt	cgaggaggtt	tcccgatatt	acctttgtgt	4500
gaaaagtctc	aatagccctt	tggtcttctg	agactgtatc	tttgatatct	ttggagttaga	4560
cgagagtgtc	gtgctccacc	atgttatcac	atcaatccac	ttgctttgaa	gacgtgggtg	4620
gaacgtcttc	tttttccacg	atgctcctcg	ttgggtgggg	tccatctttg	ggaccactgt	4680
cggcagaggg	atcttgaacg	atagcctttc	ctttatcgca	atgatggcat	ttgtagggtgc	4740
caccttcctt	ttctactgtc	cttttgatga	agtgacagat	agctgggcaa	tggaaatccga	4800
ggaggtttcc	cgatattacc	ctttgttgaa	aagcttcaat	agcccttttg	tcttctgaga	4860
ctgtatcttt	gatattcttg	gagtagacga	gagtgctcgt	ctccaccatg	ttggcaagct	4920

gctctagcca	atacgcaaac	cgctctcccc	cgcgcggttg	ccgattcatt	aatgcagctg	4980
gcacgacagg	tttcccgcact	ggaaaagcggg	cagtgagcgc	aacgcaatta	atgtgagtta	5040
gctcactcat	taggcacccc	aggctttaca	ctttatgctt	ccggctcgta	tggtgtgtgg	5100
aattgtgagc	ggataacaat	ttcacacagg	aaacagctat	gaccatgatt	acgaattcga	5160
gccttgacta	gaggggtcgac	gggtatacaga	catgataaga	tacattgatg	agtttggaca	5220
aaccacaact	agaatgcagt	gaaaaaaatg	ctttatattgt	gaaatttgtg	atgctattgc	5280
tttatttcta	accattataa	gctgcaataa	acaagttggg	gtgggacgaag	aactccagca	5340
tgagatcccc	gcgctggagg	atcatccagc	cggcgtcccc	gaaaaacgatt	ccgaagccca	5400
acctttcata	gaaggcggcg	gtggaatcga	aatctcgtag	cacgtgtcag	tcctgctcct	5460
cgccacagaa	gtgcacgcag	ttgcccggcg	ggtcgcgcag	ggcgaactcc	cgccccacg	5520
gctgctcgcc	gatctcggtc	atggccggcc	cgagagcgctc	ccggaagtcc	gtggacacga	5580
cctccgacca	ctcggcgtac	agctcgtcca	ggccgcgcac	ccacacccag	gccagggtgt	5640
tgctcggcac	cacctgggtc	tggaccgcgc	tgatgaacag	ggtcacgtcg	tcccggacca	5700
caccggcgaa	gtcgctctcc	acgaagtccc	ggaggaaccc	gagccggctcg	gtccagaact	5760
cgacgcgtcc	ggcgacgtcg	cgcgcggtga	gcaccggaac	ggcactgggtc	aacttggcca	5820
tggtaccaga	tttcgctcaa	gttagtataa	aaaagcaggc	ttcaatcctg	caggaattcg	5880
atcgacactc	togtctactc	caagaatcac	aaagatacac	tctcagaaga	ccaaagggtc	5940
attgagactt	ttcaacaaag	ggtaatatcg	ggaaacctcc	tcggattcca	ttgcccagct	6000
atctgtcact	tcatacaaaag	gacagtagaa	aaggaagggtg	gcacctacaa	atgccatcat	6060
tgcgataaag	gaaaggctat	cgttcaagat	gcctctgcgc	acagtggctc	caaagattga	6120
ccccacacca	cgaggagcat	cgtggaaaaa	gaagacgttc	caaccacgtc	ttcaaagcaa	6180
gtggattgat	gtgataacat	gggtggagcac	gacactctcg	tctactccaa	gaatatcaaa	6240
gacacagtct	cagaagacca	aagggtctatt	gagacttttc	aacaaagggt	aatatcgga	6300
aacctcctcg	gattccattg	cccagctatc	tgctacttca	tcaaaaggac	agtagaaaag	6360
gaagggtggca	cctacaaaatg	ccatcattgc	gataaaggaa	aggctatcgt	tcaagatgcc	6420
tctgcgcgaa	gtggtcccaa	agatggaccc	ccaccacga	ggagcatcgt	ggaaaaagaa	6480
gacgttccaa	ccacgtcttc	aaagcaagtg	gattgatgtg	atatctccac	tgacgttaagg	6540
gatgacgcac	aatccacta	tccttcgcaa	gaccttctc	tataaagga	agttcatttc	6600
atctggagag	gacacgctga	aatcaccagt	ctctctctac	aaatctatct	ctctcgagct	6660
ttcgagatc	cggggggggca	atgagatatg	aaaaagcctg	aactcaccgc	gacgtctgtc	6720
gagaagtttc	tgatcgaaaa	gttcgacagc	gtctccgacc	tgatgcagct	ctcggagggc	6780
gaagaatctc	gtgctttcag	cttcgatgta	ggagggcgtg	gatatgtcct	gcgggtaaat	6840
agctgcgcgc	atgggtttcta	caaagatcgt	tatgtttatc	ggcactttgc	atcgcccgcg	6900
ctcccgatcc	cggaagtgtc	tgacattggg	gagtttagcg	agagcctgac	ctattgcatc	6960
tcccgcgcgtg	cacagggtgt	cacgttgcaa	gacctgcctg	aaaccgaact	gcccgcgtgt	7020
ctacaaccgg	tcgcggaggc	tatggatgcg	atcgctgcgg	ccgatcttag	ccagacgagc	7080
gggttcggcc	cattcggacc	gcaaggaaatc	ggtcaataca	ctacatggcg	tgatttcata	7140
tgccgcatgt	ctgatcccca	tgtgtatcac	tggcaaaactg	tgatggacga	caccgtcagt	7200
gcgtccgtcg	cgcaggctct	cgatgagctg	atgctttggg	ccgaggactg	ccccgaagtc	7260
cggcacctcg	tgacgcggga	tttcggctcc	aacaatgtcc	tgacggacaa	tggccgcata	7320
acagcgggtc	ttgactggag	cgaggcgatg	ttcgggggatt	cccaatacga	ggtcgccaac	7380
atctcttctc	ggaggccgtg	gttggcttgt	atggagcagc	agacgcgcta	cttcgagcgg	7440
aggcatccgg	agcttgacag	atcgccacga	ctccgggcgt	atatgctccg	catgtgtctt	7500
gaccaactct	atcagagctt	ggttgacggc	aatctcgatg	atgcagcttg	ggcgcagggt	7560
cgatgcgacg	caatcgccg	atccggagcc	gggactgtcg	ggcgtacaca	aatcgcccg	7620
agaagcgcgg	ccgtctggac	cgatggctgt	gtagaagtac	tcgccgatag	tggaaaccga	7680
cgccccagca	ctcgtccgag	ggcaaaagaa	tagagtagat	gccgaccgga	tctgtcgatc	7740
gacaagctcg	agtttctcca	taataatgtg	tgagtagttc	ccagataaag	gaattagggt	7800
tcctataggg	tttcgctcat	gtgttgagca	tataagaaac	ccttagtatg	tatttgtatt	7860
tgtaaaatac	ttctatcaat	aaaattttcta	attcctaaaa	ccaaaatcca	gtactaaaat	7920
ccagatcccc	cgaattaatt	cggcgttaat	tcagatcaag	cttggcactg	gccgtcgttt	7980
tacaacgtcg	tgactgggaa	aaccctggcg	ttacccaact	taatcgctt	gcagcacatc	8040
cccctttcgc	cagctggcgt	aatagcgaag	aggcccgcac	cgatcgccct	tcccaacagt	8100
tgccgagcct	gaatggcgaa	tgctagagca	gcttgagctt	ggatcagatt	gtcgtttccc	8160
gccttcagtt	tggggatcct	ctagactgaa	ggcgggaaac	gacaatctga	tcatgagcgg	8220
agaattaagg	gagtcacgtt	atgacccccg	ccgatgacgc	gggacaaagcc	gttttacgtt	8280
tggaactgac	agaaccgcaa	cgttgaagga	gccactcagc	cgcggttttc	tggagtttaa	8340
tgagctaagc	acatacgtca	gaaaccatta	ttgcgcgttc	aaaagtcgcc	taaggctcact	8400
atcagctagc	aaatatttct	tgtcaaaaaa	gctccactga	cgttccataa	attccccctg	8460
gtatccaatt	agagtctcat	attcactctc	aatcccaata	atctgcaccg	gatctcgaga	8520
atcgaattcc	cgcgcccgcc	atggtagatc	tgactagtaa	aggagaagaa	cttttctactg	8580
gagttgtccc	aattcttgtt	gaatttagatg	gtgatgttaa	tgggacacaaa	ttttctgtca	8640
gtggagaggg	tgaaggtagt	gcaacatacg	gaaacttac	ccttaaat	atttgcacta	8700
ctggaaaact	acctgttcog	tggccaacac	ttgtcactac	tttctcttat	ggtgttcaat	8760
gcttttcaag	ataccagat	catatgaagc	ggcagcactt	cttcaagagc	gccatgcctg	8820
agggatacgt	ccagtagagg	accatcttct	tcaaggacga	cggaactac	aagacagctg	8880
ctgaagtcaa	gtttgagggg	gacacctctg	tcaacaggat	cgagcttaag	ggaatcgatt	8940

tcaaggagga	cggaacatc	ctcggccaca	agttggaata	caactacaac	tcccacaacg	9000
tatacatcat	ggccgacaag	caaaagaacg	gcatcaaagc	caacttcaag	acccgccaca	9060
acatcgaaga	cgccggcggtg	caactcgctg	atcattatca	acaaaataact	ccaattggcg	9120
atggccctgt	cctttttacca	gacaaccatt	acctgtccac	acaatctgcc	ctttcgaaag	9180
atcccaacga	aaagagagac	cacatggtcc	ttcttgagtt	tgtaacagct	gctgggatta	9240
cacatggcat	ggatgaacta	tacaaagcta	gccaccacca	ccaccaccac	gtgtgaattg	9300
gtgaccagct	cgaattttccc	cgatcgttca	aacatttggc	aataaaagtt	cttaagattg	9360
aatcctgttg	ccggtcttgc	gatgattatc	atataatttc	tgttgaatta	cgttaaagcat	9420
gtaataatta	acatgtaatg	catgacgtta	tttatgagat	gggtttttat	gattagagtc	9480
ccgcaattat	acattttaata	cgcatagaa	aacaaatat	agcgcgcaaa	ctaggataaa	9540
ttatcgcgcg	cgggtgcatc	tatgttacta	gatcgggaat	taaaactatca	gtgtttgaca	9600
ggatatattg	gcgggttaaac	ctaagagaaa	agagcggtta	ttagaataac	ggatatatta	9660
aaggcgctga	aaagggttat	ccgttcgtcc	atttgtatgt	gcatgccaac	cacagggttc	9720
ccctcgggat	caaaagactt	tgatccaacc	cctccgctgc	tatagtgcag	tcggcttctg	9780
acgttcagtg	cagccgtctt	ctgaaaacga	catgtcgcac	aagtcctaag	ttacgcgcga	9840
ggctgcgcgc	ctgccctttt	cctggcgctt	tcttgcgcg	tgtttttagtc	gcataaagta	9900
gaataacttg	gactagaacc	ggagacatta	gcctatgaac	aagagcgccg	ccgctggcct	9960
gctgggctat	gcccgcgtca	gcaccgacga	ccaggacttg	accaaccaac	gggcccgaact	10020
gcacgcggcc	ggctgcacca	agctgttttc	cgagaagatc	accggcacca	ggcgcgaccg	10080
cccggagctg	gccaggatgc	ttgaccacct	acgcccctgg	gacgttgtga	cagtgaccag	10140
gctagaccgc	ctggcccgca	gcaccccgca	cctactggac	attgccgagc	gcatccagga	10200
ggccggcgcg	ggcctgcgta	gcctggcaga	gccgtggggc	gacaccacca	cgccggcgccg	10260
ccgcatgggt	ttgaccgtgt	tcgcccgcac	tgccgagttc	gagcgttccc	taatcatcga	10320
ccgcaccccg	agcggcgcg	agcccgccaa	ggcccagggc	gtgaagtttg	gccccgcgcc	10380
taccctcacc	ccggcacaga	tcgcgcacgc	ccgcgagctg	atcgaccagg	aaggccgcac	10440
cgtgaaagag	gcggctgcac	tgettggcgt	gcattcgctc	accctgtacc	gcgcacttga	10500
gcgcagcgag	gaagtgcgcg	ccaccgaggg	caggcgccgc	ggtgccttcc	gtgaggacgc	10560
attgaccgag	gccgacgccc	tgccgcccgc	cgagaatgaa	cgccaagagg	aacaagcatg	10620
aaaccgcacc	aggacggcca	ggacgaaccg	tttttcatta	ccgaagagat	cgaggcggag	10680
atgatcgcgg	ccgggtacgt	gttcgagccg	cccgcgcacg	tctcaaccgt	gcggctgcac	10740
gaaatcctgg	ccggtttgtc	tgatgccaa	ctggcgccct	ggccggccag	cttggccgct	10800
gaagaaaccg	agcggccgcg	tctaaaaagg	tgatgtgtat	ttgagtaaaa	cagettgcgt	10860
catgcggtcg	ctgcgtatat	gatgcgatga	gtaaaataaac	aaatacgcga	ggggaacgca	10920
tgaaggttat	cgctgtactt	aaccagaaag	gcgggtcagg	caagacgacc	atcgcaaccc	10980
atctagcccc	cgccctgcaa	ctcgccgggg	ccgatgttct	gttagtcgat	tccgatcccc	11040
aggcgagtg	ccggcattgg	gcggcccggt	gggaagatca	accgctaacc	gttgtcgga	11100
tcgaccgccc	gacgattgac	cgcgacgtga	aggccatcgg	ccggcgcgac	ttcgtagtga	11160
tcgaccggag	gccccaggcg	gcggacttgg	ctgtgtccgc	gatcaaggca	gccgacttcg	11220
tgctgattcc	ggtgcagcca	agcccttacg	acatatgggc	caccgcgcac	ctggtggagc	11280
tggttaagca	gcgcattgag	gtcacggatg	gaaggctaca	agcggccttt	gtcgtgtcgc	11340
ggcgcatcaa	aggcacgcgc	atcgccgggt	agggtgccga	ggcgctggcc	gggtacgagc	11400
tgcccattct	tgagtcctgt	atcacgcagc	gcgtgagcta	cccaggcact	gccgcgcgcg	11460
gcacaaccgt	tcttgaatca	gaacccgagg	gcgacgctgc	ccgcgaggtc	caggcgctgg	11520
ccgctgaaat	taaatcaaaa	ctcatttgag	ttaatgaggt	aaagagaaaa	tgagcaaaa	11580
cacaaacacg	ctaagtgcgc	gcggtccgag	cgcacgcagc	agcaaggctg	caacgttggc	11640
cagcctggca	gacacgccag	ccatgaagcg	gggtcaactt	cagttgccgg	cggaggatca	11700
caccaagctg	aagatgtacg	cggtacgcga	aggcaagacc	attaccgagc	tgctatctga	11760
atacatcgcg	cagctaccag	agtaaatgag	caaatagaata	aatgagtaga	tgaattttag	11820
cggctaaagg	aggcggcatg	gaaaatcaag	aacaaccagg	caccgacgcc	gtggaatgcc	11880
ccatgtgtgg	aggaacgggc	ggttggccag	gcgtaagcgg	ctgggttgtc	tgccggccct	11940
gcaatggcac	tggaaacccc	aagcccagg	aatcggcggt	acggtcgcaa	accatccggc	12000
ccggtacaaa	tcggcgcgcc	gctgggtgat	gacctgggtg	agaagtga	ggccgcgcag	12060
gccgcccagc	ggcaacgcac	cgaggcgaaa	gcacgccccg	gtgaatcgtg	gcaagcggcc	12120
gctgatcgaa	tcgcgcaaga	atcccggcaa	ccgcgcgcag	ccggtgcgcc	gtcgattagg	12180
aagccgcccc	agggcgacga	gcaaccagat	tttttcgttc	cgatgctcta	tgacgtgggc	12240
acccgcgata	gtcgagcat	catggacgtg	cccggttttc	gtctgtcgaa	gcgtgaccca	12300
cgagctggcg	aggttagccg	ctacgagctt	gcagacgggc	acgtagaggt	ttccgcaggg	12360
ccggccggca	tggccagtgt	gtgggattac	gacctggtac	tgatggcggt	ttcccatcta	12420
accgacatcca	tgaaccgata	ccgggaaggg	aaggagagaca	agcccggccg	cgtgttccgt	12480
ccacacgttg	cggacgtact	caagttctgc	cgggcagccg	atggcgga	gcagaaagac	12540
gacctggtag	aaacctgcac	tcggttaaac	accacgcacg	ttgccatgca	gc	12592

<210> 96
 <211> 3357
 <212> DNA
 <213> Artificial Sequence

<220>

<223> pGEMEasyNOS Plasmid

<400> 96

tatcactagt	gaatttcgagg	ccgcctgcag	gtcgaccata	tgggagagct	cccaacgcgt	60
tggatgcata	gcttgagtat	tctatagtgt	cacctaaata	gcttggcgta	atcatgggtca	120
tagctgtttc	ctgtgtgaaa	ttgttatccg	ctcacaaattc	cacacaacat	acgagccgga	180
agcataaagt	gtaaagcctg	gggtgcctaa	tgagtgcgct	aactcacatt	aattgcggtg	240
cgctcactgc	ccgctttcca	gtcgggaaac	ctgtcgtgcc	agctgcatta	atgaatcggc	300
caacgcgcgg	ggagaggcgg	tttgcgattt	gggcgctctt	ccgcttccctc	gctcactgac	360
tcgctgcgct	cggctcgttcg	gctgcggcga	gcggtatcag	ctcactcaaaa	ggcggtaata	420
cgggttatcca	cagaatcagg	ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	480
aagggccagg	accgtaaaaa	ggccgcgttg	ctggcggtttt	tccataggct	ccgccccctt	540
gacgagcatc	acaaaaaatcg	acgctcaagt	cagaggtggc	gaaacccgac	aggactataa	600
agataccagg	cgtttccccc	tggaaagctcc	ctcgtgcgct	ctcctgttcc	gacctgtccg	660
cttaccggat	acctgtccgc	ctttctccct	tccgggaagcg	tggcgctttc	tcatagctca	720
cgctgtagggt	atctcagttc	ggtgtaggtc	gttcgctcca	agctgggctg	tgtgcacgaa	780
ccccccggttc	agcccgaccg	ctgcgcctta	tccggttaact	atcgtcttga	gtccaacccg	840
gtaagacacg	acttatcgcc	actggcagca	gccattggta	acaggattag	cagagcggag	900
tatgtaggcg	gttcttgaga	gttcttgaa	tgggtggccta	actacggcta	actagaaga	960
acagtatttg	gtatctgcgc	tctgctgaag	ccagttacct	tccgaaaaag	agttggtagc	1020
tcttgatccg	gcaaaacaaac	caccgctgggt	agcgggtgggt	tttttggttg	caagcagcag	1080
attacgcgca	gaaaaaaagg	atctcaagaa	gatcctttga	tcttttctac	ggggtctgac	1140
gctcagtgga	acgaaaaactc	acgttaagggt	attttggtca	tgagattatc	aaaaaggatc	1200
ttcacctaga	tcctttttaa	ttaaaaatga	agttttaaat	caatctaaag	tatatatgag	1260
taaacctgggt	ctgacagtta	ccaatgctta	atcagtgagg	cacctatctc	agcgatctgt	1320
ctatttctgt	catccatagt	tgccctgactc	cccgctcgtgt	agataactac	gatacgggag	1380
ggcttaccat	ctggcccccag	tgctgcaatg	ataccgcgag	acccacgctc	accggctcca	1440
gatttatcag	caataaacca	gccagccgga	agggccgagc	gcagaagtgg	tcctgcaact	1500
ttatccgcct	ccatccagtc	tattaatgtgt	tgccgggaaag	ctagagtaag	tagttcgcca	1560
gttaaatagtt	tgcgcaacgt	tggtgccatt	gctacaggca	tccgtggtgtc	acgctcgtcg	1620
tttggtatgg	cttcattcag	ctccggttcc	caacgatcaa	ggcgagttac	atgatcccc	1680
atgttgtgca	aaaaagcgggt	tagctccttc	ggtcctccga	tccgttgtcag	aagtaagttg	1740
gccgcagtggt	tatcactcat	ggttatggca	gcactgcata	attctcttac	tgtcatgcca	1800
tccgtaagat	gcttttctgt	gactggtgag	tactcaacca	agtcattctg	agaatagtg	1860
atgcggcgac	cagagtgtctc	ttgcccggtg	tcaatacggg	ataataccgc	gccacatagc	1920
agaactttaa	aagtgtctcat	cattggaaaa	cgttcttcgg	ggcgaaaaact	ctcaaggatc	1980
ttaccgctgt	tgagatccag	ttcgatgtaa	cccactcgtg	cacccaactg	atcttcagca	2040
tcctttactt	tcaccagcgt	ttctgggtga	gcaaaaaacag	gaaggcaaaa	tgccgcaaaa	2100
aagggaataa	ggcgacacag	gaaatgttga	atactcatat	tcttcctttt	tcaatattat	2160
tgaagcattt	atcaggggta	ttgtctcatg	agcggataca	tatttgaatg	tatttagaaa	2220
aataaaacaaa	taggggttcc	gcgcacattt	ccccgaaaag	tgccacctga	tgccgtgtga	2280
aataccgcac	agatgcgtaa	ggagaaaaata	ccgcatcagg	aaattgtaag	cggttaattt	2340
ttgttaaaat	tcgcgttaaa	tttttggtta	atcagctcat	tttttaacca	ataggccgaa	2400
atcggcaaaa	tcctttataa	atcaaaagaa	tagaccgaga	taggggttgag	tgttgttcca	2460
gtttggaaca	agagtccact	attaaagaac	gtggactcca	acgtcaaagg	gcgaaaaaac	2520
gtctatcagg	gcgatggccc	actacgtgaa	ccatcacctc	aatcaagttt	tttgggggtc	2580
aggtgccgta	aagcactaaa	tcggaaacct	aaaggagacc	cccgatattag	agcttgacgg	2640
ggaaagccgg	cgaacgtggc	gagaaaggaa	gggaaggaa	cgaaaggagc	gggcgctagg	2700
gcgctggcaa	gtgtagcggg	cacgctgcgc	gtaaccacca	cacccgccgc	gcttaatgcg	2760
ccgctacagg	gcgcgtccat	tcgccattca	ggctgcgcaa	ctgttgggaa	gggcgatcgg	2820
tgccggcctc	ttcgctatta	cgccagctgg	cgaaaggggg	atgtgctgca	aggcgattaa	2880
gttgggtaac	gccaggggtt	tcccagtcac	gacgttgtta	aacgacggcc	agtgaattgt	2940
aatacgactc	actatagggc	gaattggggc	cgacgtcgca	tgctcccggc	cgccatggcg	3000
gccgcgggaa	ttcgattctc	gagatccggg	gcagattatt	tggattgaga	gtgaatatga	3060
gactctaatt	ggataccgag	gggaatttat	ggaacgtcag	tggagcattt	ttgacaagaa	3120
atatttgcta	ctgatagtg	accttagggc	acttttgaac	gcgcaataat	ggtttctgac	3180
gtatgtgctt	agctcattaa	actccagaaa	cccgcggctg	agtggtcctt	tcaacgttgc	3240
ggttctgtca	gttccaaacg	taaaacgggt	tgtcccgcgt	catcggcggg	ggtcataacg	3300
tgactccctt	aattctccgc	tcatgatcag	attgtcgttt	cccgccttca	gtctaga	3357

<210> 97

<211> 10122

<212> DNA

<213> Artificial Sequence

<220>

<223> p1302NOS Plasmid

<400> 97

catggttagat	ctgactagta	aaggagaaga	acttttctact	ggagttgtcc	caattcttgt	60
tgaattagat	gggtgatgta	atggggcacia	attttctgtc	agtggagagg	gtgaaggtga	120
tgcaacatac	ggaaaactta	cccttaaatt	tatttgcact	actggaaaac	tacctgttcc	180
gtggccaaca	cttgtcacta	ctttctctta	tgggtgttcaa	tgcttttcaa	gatacccaga	240
tcatatgaag	cggcacgact	tcttcaagag	cgccatgcct	gagggatacg	tgaggagag	300
gaccatcttc	ttcaaggacg	acgggaacta	caagacacgt	gctgaagtca	agtttgaggg	360
agacaccctc	gtcaacagga	tcgagcttaa	gggaatcgat	ttcaaggagg	acggaaacat	420
cctcggtccac	aagttggaat	acaactacaa	ctcccacaa	gtatacatca	tgcccgacaa	480
gcaaaagaac	ggcatcaaag	ccaacttcaa	gacccgccac	aacatcgaag	acggcgccgt	540
gcaactcgct	gatcattatc	aacaaaatac	tcgaattggc	gatggccctg	tccttttacc	600
agacaacctat	tacctgtcca	cacaatctgc	cctttcgaaa	gatcccaacg	aaaagagaga	660
ccacatggtc	cttcttgagt	ttgtaacagc	tgctgggatt	acacatggca	tggatgaact	720
atacaagct	agccaccacc	accaccacca	cgtgtgaatt	ggtagaccag	tcgaatttcc	780
ccgatcgctt	aaacttttgg	taataaagtt	tcttaagatt	gaatcctgtt	gccgtctgt	840
cgatgattat	catataattt	ctgttgaatt	acgttaagca	tgtaataatt	aacatgtaat	900
gcgatcgctt	atttatgaga	tgggtttttaa	tgattagagt	cccgcaatta	tacatttaac	960
acgcgataga	aaacaaaata	tagcgcgcaa	actaggataa	attatcgccg	gcggtgtcat	1020
ctatgttact	agatcgggaa	ttaaactatc	agtgtttgac	aggatatatt	ggcgggtaaa	1080
cctaagagaa	aagagcggtt	attagaataa	cggatatatta	aaagggcggt	aaaaggttta	1140
tccgttctgc	catttctgtg	tgcatgccaa	ccacagggtt	cccctcggga	tcaaagtact	1200
ttgatccaac	ccctccgctg	ctatagtcca	gtcgggttct	gacgttcagt	gcagccgtct	1260
tctgaaaacg	acatgtcgca	caagtcctaa	gttacgcgac	aggctgccgc	cctgcccttt	1320
tcctggcggt	ttcttctgcg	gtgttttagt	cgcataaagt	agaatacttg	cgactagaac	1380
cggagacatt	acgccatgaa	caagagcgcc	gcccgtggcc	tgctgggcta	tgcccgcgct	1440
agcaccgacg	accaggactt	gaccaaccaa	cgggccgaac	tgacgcgggc	cggtgcacc	1500
aagctgtttt	ccgagaagat	caccggcacc	aggcgcgacc	gcccggagct	ggccaggatg	1560
cttgaccacc	tacgccctgg	cgacgtttgt	acagtaccca	ggctagaccg	cctggcccgc	1620
agcaccgctg	acctaactgga	catttgcgag	cgcattccagg	aggccggcgc	gggcctgcgt	1680
agcctggcga	agccgtgggc	cgacaccacc	acgcgcggcg	gcccgcattg	gttgaccgtg	1740
ttcgccggta	ttgcctaggt	cgagcgttcc	ctaattcatcg	accgcaccgc	gagcggggcg	1800
gagggcgcca	aggcccgagg	cgtgaagttt	ggcccccgcc	ctaccctcac	cccggcacag	1860
atcgcgacg	cccgcgagct	gatcgaccag	gaaggccgca	ccgtgaaaga	ggcggttgca	1920
ctgcttggcg	tgcatcgctc	gacctgttac	cgcgacttgc	agcgacgcca	ggaagtgcag	1980
cccaccgagg	ccaggcggtg	cggtgccttc	cgtgaggacg	cattgaccga	ggccgacgcc	2040
ctggcgggcg	ccgagaatga	acgccaaagag	gaacaagcat	gaaaccgcac	caggacggcc	2100
aggacgaacc	gttttttcat	acggaagaga	tcgagcgcca	gatgatcgcg	gcccgggtacg	2160
tggtcgagcc	gcccgcgcac	gtctcaaccg	tgccggtgca	tgaaatcctg	gcccgtttgt	2220
ctgatgccaa	gctggcgggc	tgcccgggcca	gcttggccgc	tgaaagaaac	gagcgccgcg	2280
gtctaaaaag	gtgatgtgta	tttgagttaa	acagcttgcc	tcattgcggtc	gctgcgtata	2340
tgatgcgatg	agtaaataaa	caaatacgca	aggggaacgc	atgaaggtta	tcgctgtact	2400
taaccagaaa	ggcgggtcag	gcaagacgac	catcgcaacc	catctagccc	gcgccttgca	2460
actcgccggg	gccgatgttc	tgtagtcgca	ttccgatccc	cagggcagtg	cccgcgattg	2520
ggcgcccggt	cgggaagatc	aaccgctaac	cgttgtcgcc	atcgaccgcc	cgacgattga	2580
ccgcgacgtg	aaggccatcg	gccggcgcca	cttcgtagtg	atcgacggag	cgccccagcg	2640
ggcggaactg	gctgtgtccg	cgatcaaggc	agccgacttc	gtgctgattc	cggtgcagcc	2700
aagcccttac	gacatatggg	ccaccgcgca	cctggtggag	ctggttaagc	agcgcatgta	2760
ggtcacggat	ggaaggctac	aagcggcctt	tgtcgtgtcg	cgggcgatca	aaggcacgcg	2820
catcgccggg	gaggttgccg	aggcgctggc	cgggtacgag	ctgcccattc	ttgagtcgcc	2880
tatcacgcag	cgcgtgagct	acccaggcac	tgccgcggcc	ggcacaaccg	ttcttgaatc	2940
agaacccgag	ggcgacgctg	cccgcgaggt	ccaggcgctg	gccgctgaaa	ttaaatcaaa	3000
actcatttga	gttaatgagg	taaagagaaa	atgagcaaaa	gcacaaacac	gctaagtgc	3060
ggccgtccga	gcgcacgcag	cagcaaggct	gcaacgttgg	ccagcctggc	agacacgcca	3120
gccaatgaag	gggtcaactt	tcagtgtccg	gcggaggatc	acaccaagct	gaagatgtac	3180
gcggtagccc	aaggcaagac	cattaccgag	ctgctatctg	aatacatcgc	gcagctacca	3240
gagtaaatga	gcaaatgaat	aaatgagtag	atgaatttta	gcggctaaag	gaggcgccat	3300
ggaaaaatcaa	gaacaaccag	gcaccgacgc	cgtggaatgc	cccatgtgtg	gaggaacggg	3360
cgggttgccc	ggcgtaacgc	gctgggttgt	ctgcggggcc	tgcaatggca	ctggacccc	3420
caagcccgag	gaatcgccgt	gacggctcgca	aaccatccgg	cccgggtacaa	atcgccgcgg	3480
cgctgggtga	tgacctgggt	gagaagttga	agggccgcga	ggccgcccag	cgcccaacga	3540
tcgaggcgaga	agcacgcccc	ggtgaatcgt	ggcaagcgcc	cgctgatcga	atccgcaagg	3600
aatcccggca	accgcccggca	gccgggtgcgc	cgtcgattag	gaagccgccc	aagggcgacg	3660
agcaaccaga	ttttttcggt	ccgatgctct	atgacgtggg	cacccgcgat	agtcgcagca	3720
tcatggacgt	ggccggtttt	cgtctgtcga	agcgtgaccc	acgagctggc	gaggtgatcc	3780
gctacgagct	tccagacggg	cacgtagagg	tttccgcagg	gccggccggc	atggccagtg	3840

tgtgggatta	cgacctggta	ctgatggcgg	tttcccatct	aaccgaatcc	atgaaccgat	3900
accgggaagg	gaagggagac	aagcccggcc	gcgtgttccg	tccacacgtt	gcggacgtac	3960
tcaagttctg	ccggcgagcc	gatggcggaa	agcagaaaga	cgacctggta	gaaacctgca	4020
ttcgggttaa	caccacgcac	gttgccatgc	agcgtacgaa	gaaggccaag	aacggccgcc	4080
tgggtgacggt	atccgagggg	gaagccttga	ttagccgcta	caagatcgta	aagagcgaaa	4140
ccgggcggcc	ggagtacatc	gagatcgagc	tagctgattg	gatgtaccgc	gagatcacag	4200
aaggcaagaa	cccgagcgtg	ctgacgggtt	accccgatta	cttttttgatc	gatcccgcca	4260
tcgcccggtt	tctctaccgc	ctggcacgcc	gcgcgcgagg	caaggcagaa	gccagatggg	4320
tgttcaagac	gatctacgaa	cgcagtggca	gcgcgcggaga	ggttcaagaag	ttctgtttca	4380
ccgtgcgcaa	gctgatcggg	tcaaattgacc	tgcgcggagta	cgatttgaag	gaggaggcgg	4440
ggcaggctgg	cccgatccta	gtcatgcgct	accgcaacct	gatcgagggg	gaagcatccg	4500
ccggttccta	atgtacggag	cagatgctag	ggcaaatgct	cctagcaggg	gaaaaagggtc	4560
gaaaagggtc	ctttcctgtg	gatagcacgt	acattgggaa	cccaaaggccg	tacattggga	4620
accggaaccc	gtacattggg	aacccaaagc	cgtaattggg	gaaccggtca	cacatgtaac	4680
tgactgatat	aaaagagaaa	aaaggcgatt	tttccgccta	aaactcttta	aaacttatta	4740
aaactcttaa	aacccgcctg	gcctgtgcat	aactgtctgg	ccagcgcaca	gccgaagagc	4800
tgcaaaaagc	gcctaccctt	cggtcgctgc	ccccgcgctc	ccccgcgctc	tcgcgtcgcc	4860
ctatcgcgcc	cgctggccgc	tcaaaaatgg	ctggcctacg	gccaggcaat	ctaccagggc	4920
gcgggacaagc	cgcgccgctg	ccactcgacc	gcgcgcgccc	acatcaaggc	accctgcctc	4980
gcgcggtttc	gtgatgcagg	tgaaaaacct	tgacacatgc	agctcccggg	gacggtcaca	5040
gcttgtctgt	aagcggatgc	cgggagcaga	caagcccgtc	agggcgcgct	agcgggtgtt	5100
ggcgggtgtc	ggggcgcgag	catgaccacg	tcacgtagcg	atagcggagt	gtatactggc	5160
ttactatgc	ggcatcagag	cagattgtac	tgagagtgcg	ccatatgcgg	tgtgaaatac	5220
cgcacagatg	cgtaaggaga	aaataccgca	tcagcgctc	ttccgcttcc	tcgctcactg	5280
actcgctgcg	ctcggtcggt	cggtcgcgcc	gagcgggtatc	agctcactca	aaggcggtaa	5340
tacggttatc	cacagaatca	ggggataacg	caggaaagaa	catgtgagca	aaaggccagc	5400
aaaaggccag	gaaccgtaaa	aaggccgctg	tgctggcggt	tttccatagg	ctccgcccc	5460
ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagagggtg	gcgaaacccg	acaggactat	5520
aaagatacca	ggcggtttccc	cctggaagct	ccctcgctgc	ctctcctgtt	ccgacctgc	5580
cgcttaccgg	atacctgtcc	gcctttctcc	cttcgggaag	cgctggcgctt	tctcatagct	5640
cacgctgtag	gtatctcagt	tcggtgtagg	tcgttcgctc	caagctgggc	tgtgtgcacg	5700
aacccccctg	tcagcccgac	cgctgcgcct	tatccggtaa	ctatcgctct	gagtccaacc	5760
cggttaagaca	cgacttatcg	ccactggcag	cagccactgg	taacaggatt	agcagagcga	5820
ggtagtagg	cggtgctaca	gagttcttga	agtgggtggc	taactacggc	tacactagaa	5880
ggacagattt	tggtatctgc	gctctgctga	agccagttac	cttcggaaaa	agagttggta	5940
gctcttgatc	gggcaaacaa	accaccgctg	gtagcgggtg	tttttttggt	tgcaagcagc	6000
agattacgcg	cagaaaaaaa	ggatctcaag	aagatccttt	gatcttttct	acgggggtctg	6060
acgctcagtg	gaacgaaaaa	tcacgttaag	ggatttttgt	catgcattct	aggtactaaa	6120
acaattcatc	cagtaaaaata	taataatttta	ttttctccca	atcagggttg	atccccagta	6180
agtcaaaaaa	tagctcgaca	tactgttctt	ccccgatatc	ctccctgatc	gaccggacgc	6240
agaaggcaat	gtcataccac	ttgtccgccc	tgcgccttct	cccaagatca	ataaagccac	6300
ttactttggc	atctttcaca	aagatgttgc	tgtctcccag	gtcgcggtgg	gaaaagacaa	6360
gttctctctc	gggcttttcc	gtctttaaaa	aatcatacag	ctcgcgcgga	tctttaaatg	6420
gagtgctctc	ttccaggttt	tcgcaatcca	catcgccag	atcgttattc	agtaagtaat	6480
ccaattcggc	taagcggctg	tctaagctat	tcgtataggg	acaatccgat	atgtcgatgg	6540
agtgaagag	cctgatgcac	tcgcataca	gctcgataat	cttttcaggg	ctttgttcat	6600
cttcatactc	ttccgagcaa	aggacgccat	cgccctcact	catgagcaga	ttgtctccagc	6660
catcatgccc	ttcaaagtgc	aggacctttg	gaacaggcag	ctttccttcc	agccatagca	6720
tcattgtcctt	ttcccgcttc	acatcatagg	tggtcccttt	ataccggctg	tcgctcattt	6780
ttaaatatag	gttttccatt	tctcccacca	gcttatatac	cttagcagga	gacattcctt	6840
ccgtatcttt	tacgcagcgg	tatttttcga	tcagtttttt	caattccggt	gatattctca	6900
tttttagccat	ttattatttt	cttcctcttt	tctacagtat	ttaaagatac	cccaagaagc	6960
taattataac	aagacgaact	ccaattcact	gttcccttgca	ttctaaaaacc	ttaaatacca	7020
gaaaacagct	ttttcaaagt	tgttttcaaa	gttggcgctat	aacatagtat	cgacggagcc	7080
gattttgaaa	ccgcgggtgat	cacaggcagc	aacgctctgt	catcggtaca	atcaacatgc	7140
taccctccgc	gagatcatcc	gtgtttcaaa	cccgccagct	tagttgccgt	tcttccgaat	7200
agcatcggtg	acatgagcaa	agtctgccgc	cttacaacgg	ctctcccgtc	gacgccgtcc	7260
cggactgatg	ggctgectgt	atcgagtggg	gatttttgtgc	cgagctgccg	gtcgggggagc	7320
tgttggctgg	ctgggtggcag	gatataattgt	gggtgaaaca	aattgacgct	tagacaactt	7380
aataacacat	tgccagcgtt	tttaattgat	tgaattaaacg	ccgaattaat	tcgggggcatc	7440
tggatttttag	tactggattt	tggtttttagg	aattagaaat	tttattgata	gaagtatttt	7500
acaaatacaa	atacatatac	aggggtttctt	atatgtctcaa	cacatgagcg	aaaccctata	7560
gggaacccata	ttcccttatac	tgggaactac	tcacacatta	ttatggagaa	actcgagctt	7620
gtcgatcgac	agatccgggtc	ggcatctact	ctattttcttt	gccctcggac	gagtgctggg	7680
cgctcggttt	ccactcatcg	cgagtacttc	tacacagcca	tcgggtccaga	cgcccgcgct	7740
tctcggggct	atttgtgtac	gcccagacgt	cccggtcccg	gatcggaaga	ttgcgtcgca	7800
tcgaccctgc	gcccgaagctg	catcatcgaa	attgcccgtca	accaagctct	gatagagttg	7860


```

gtcaagacca atgcggagca tatacgcgcc gagtcgtggc gatcctgcaa gctccggatg 7920
cctccgctcg aagtagcgcg tctgctgctc catacaagcc aaccacggcc tccagaagaa 7980
gatgttggcg acctcgatatt gggaatcccc gaacatcgcc tcgctccagt caatgaccgc 8040
tgttatgcgg ccattgtccg tcaggacatt gttggagccg aaatccgcgt gcacgaggtg 8100
ccggacttcg gggcagtcct cggcccaaaag catcagctca tcgagagcct gcgcgacgga 8160
cgactgacg gtgtcgtcca tcacagtttg ccagtgatac acatggggat cagcaatcgc 8220
gcatatgaaa tcacgccatg tagtgtattg accgattcct tgcggtccga atgggccgaa 8280
cccgtcgtc tggctaagat cggcgcgacg gatcgcatcc atagcctccg cgaccggttg 8340
tagaacagcg ggcagttcgg tttcaggcag gtcttgcaac gtgacaccct gtgcacggcg 8400
ggagatgcaa taggtcaggc tctcgctaaa ctccccaatg tcaagcactt ccggaatcgg 8460
gagcgcgccc gatgcaaaagt gccgataaac ataacgatct ttgtagaaac catcggcgca 8520
gctattttacc cgcaggacat atccacgccc tcctacatcg aagctgaaaag caccgagattc 8580
ttcgccctcc gagagctgca tcaggtcggg gacgctgtcg aacttttcga tcagaaactt 8640
ctcgacagac gtccggttga gttcaggctt tttcatactc cattgcccc ccggatctgc 8700
gaaagctcga gagagataga tttgtagaga gagactgggtg atttcagcgt gtccctctcca 8760
aatgaaatga acttccttat atagaggaag gtcttgcgaa ggatagtggg attgtgcgtc 8820
atcccttacg tcagtgaggaga tatcacatca atccacttgc tttgaagacg tgggtggaac 8880
gtctctcttt tccacgatgc tcctcgtagg tgggggtcca tctttgggac cactgtcggc 8940
agaggcatct tgaacgatag cctttccttt tgatgaagtg acagatagct gggcaatgga atccgaggag 9000
ttccttttct actgtccttt gttgaaaagt ctcaatagcc ctttggctct ctgagactgt 9060
atctttgata ttcttggagt agacgagagt gtctgtctcc accatgttat cacatcaatc 9120
cacttgcttt gaagacgtgg ttggaacgct tctcttttcc acgatgctcc tcgtgggtgg 9180
gggtccatct ttgggaccac tgtcggcaga ggcattctga acgatagcct ttcctttatc 9240
gcaatgatgg catttgtagg tgccaccttc cttttctact gtccctttga tgaagtgaca 9300
gatagctggg caatggaatc cgaggaggtt tcccgatatt accctttgtt gaaaagtctc 9360
aatagccott tgggtctctg agactgtatc tttgatattc ttggagtga cgagagtgtc 9420
gtgtccacc atgttggcaa gctgtcttag ccaatacgca aaccgcctct ccccgcgctg 9480
tggccgattc caatggaatc ctggcacgac aggtttcccg actggaagc cccaggtctt acactttatg 9540
cgcaacgcaa ttaatgtgag ttagctcact cattagggac agcggataac aatttcacac aggaaacagc 9600
cttcgggtc gtatgttgtg tgggaattgt agcggataac cctctagact gaaggcggga 9660
tatgaccatg attacgaatt cgagctcggg agggagtcac gttatgaccc ccgccgatga 9720
aacgacaatc tgatcatgag cggagaatta agggagtcac caacgttgaa ggagccactc 9780
cgcgggacaa ggcgttttac gtttggaaat gacagaaccg taagaaacca ttattgcgcg 9840
agcgcgggtt ttctggagtt taatgagcta agcacatacg tcttgtaaaa aatgtctccac 9900
ttcaaaagtc gcctaagggt actatcagct agcaaataat tcttgtaaaa aatgtctccac 10020
tgacgttcca taaattcccc tcggtatcca attagagtct catattcact ctcaatccaa 10080
ataatctgca ccggatctcg agaatcgaat tcccgcggcc gc 10122

```

<210> 98
 <211> 621
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> N. tabacum rDNA intergenic spacer (IGS) sequence

<300>
 <308> Genbank #Y08422
 <309> 1997-10-31

```

<400> 98
gtgctagcca atgtttaaca agatgtcaag cacaatgaat gttggtgggt ggtgggtcgtg 60
gctggcggtg gtggaaaatt gcggtgggtc gagcggtagt gatcggcgat ggttgggtgt 120
tgcagcggtg tttgatatcg gaatcactta tgggtgggtgt cacaatggag gtgcgtcatg 180
gttatgggtg gttggtcatc tatatatttt tataataata ttaagtattt tacctatttt 240
ttacatatatt tttattaaat ttatgcattg tttgtatttt taaatagttt ttatcgtaact 300
tgttttataa aatatatttt tattttatgt gtttatattt tacttgatgt attggaaatt 360
ttctccattg ttttttctat attttataata attttcctat ttttttttgt tttattatgt 420
attttttcgt tttataataa atattttatta aaaaaatat tatttttgta aaatatatca 480
tttacaatgt ttaaaagtca tttgtgaata tattagctaa gttgtacttc tttttgtgca 540
tttgggtgtg tacatgtcta ttatgattct ctggccaaaa catgtctact cctgtcactt 600
gggttttttt ttttaagaca t

```

<210> 99
 <211> 25
 <212> DNA

<213> Artificial Sequence

<220>

<223> NTIGS-F1 Primer

<400> 99

gtgctagcca atgtttaaca agatg 25

<210> 100

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> NTIGS-R1 Primer

<400> 100

atgtcttaaa aaaaaaaacc caagtgac 28

<210> 101

<211> 233

<212> DNA

<213> Mus Musculus

<300>

<308> Genbank #V00846

<309> 1989-07-06

<400> 101

gacctggaat atggcgagaa aactgaaaat cacggaaaat gagaaataca cacttttagga 60
cgtgaaatat ggcgaggaaa actgaaaaag gtggaaaatt tagaaatgtc cactgttagga 120
cgtggaatat ggcaagaaaa ctgaaaatca tggaaaatga gaaacatcca cttgacgact 180
tgaaaaatga cgaaatcact aaaaaacgtg aaaaatgaga aatgcacact gaa 233

<210> 102

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> MSAT-F1 Primer

<400> 102

aataccgcgg aagcttgacc tggaatatcg c 31

<210> 103

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> MSAT-R1 Primer

<400> 103

ataaccgcgg agtccttcag tgtgcat 27

<210> 104

<211> 277

<212> DNA

<213> Artificial Sequence

<220>

<223> Nopaline Synthase Promoter Sequence

<300>

<308> Genbank #U09365

<309> 1997-10-17

```
<400> 104
gagctcgaat ttccccgatac gttcaaacaat ttggcaataa agtttcttaa gattgaatcc 60
tggtgccggt cttgcgatga ttatcatata atttctgttg aattacgtta agcatgtaat 120
aattaacatg taatgcatga cgttatttat gagatgggtt tttatgatta ggtcccgca 180
attatacatt taatacgcga tagaaaacaa aatatagcgc gcaaactagg ataaattatc 240
gcgcgcggtg tcacttatgt tactagatcg ggaattc 277
```

```
<210> 105
<211> 1812
<212> DNA
<213> Escherichia coli
```

```
<220>
<221> CDS
<222> (1)...(1812)
<223> Beta-Glucuronidase
```

```
<300>
<308> Genbank #S69414
<309> 1994-09-23
```

```
<400> 105
atg tta cgt cct gta gaa acc cca acc cgt gaa atc aaa aaa ctc gac 48
Met Leu Arg Pro Val Glu Thr Pro Thr Arg Glu Ile Lys Lys Leu Asp
1 5 10 15

ggc ctg tgg gca ttc agt ctg gat cgc gaa aac tgt gga att gat cag 96
Gly Leu Trp Ala Phe Ser Leu Asp Arg Glu Asn Cys Gly Ile Asp Gln
20 25 30

cgt tgg tgg gaa agc gcg tta caa gaa agc cgg gca att gct gtg cca 144
Arg Trp Trp Glu Ser Ala Leu Gln Glu Ser Arg Ala Ile Ala Val Pro
35 40 45

ggc agt ttt aac gat cag ttc gcc gat gca gat att cgt aat tat gcg 192
Gly Ser Phe Asn Asp Gln Phe Ala Asp Ala Asp Ile Arg Asn Tyr Ala
50 55 60

ggc aac gtc tgg tat cag cgc gaa gtc ttt ata ccg aaa ggt tgg gca 240
Gly Asn Val Trp Tyr Gln Arg Glu Val Phe Ile Pro Lys Gly Trp Ala
65 70 75 80

ggc cag cgt atc gtg ctg cgt ttc gat gcg gtc act cat tac ggc aaa 288
Gly Gln Arg Ile Val Leu Arg Phe Asp Ala Val Thr His Tyr Gly Lys
85 90 95

gtg tgg gtc aat aat cag gaa gtg atg gag cat cag ggc ggc tat acg 336
Val Trp Val Asn Asn Gln Glu Val Met Glu His Gln Gly Gly Tyr Thr
100 105 110

cca ttt gaa gcc gat gtc acg ccg tat gtt att gcc ggg aaa agt gta 384
Pro Phe Glu Ala Asp Val Thr Pro Tyr Val Ile Ala Gly Lys Ser Val
115 120 125

cgt atc acc gtt tgt gtg aac aac gaa ctg aac tgg cag act atc ccg 432
Arg Ile Thr Val Cys Val Asn Asn Glu Leu Asn Trp Gln Thr Ile Pro
130 135 140

ccg gga atg gtg att acc gac gaa aac ggc aag aaa aag cag tct tac 480
Pro Gly Met Val Ile Thr Asp Glu Asn Gly Lys Lys Lys Gln Ser Tyr
145 150 155 160

ttc cat gat ttc ttt aac tat gcc gga atc cat cgc agc gta atg ctc 528
Phe His Asp Phe Phe Asn Tyr Ala Gly Ile His Arg Ser Val Met Leu
165 170 175

tac acc acg ccg aac acc tgg gtg gac gat atc acc gtg gtg acg cat 576
```

Tyr	Thr	Thr	Pro	Asn	Thr	Trp	Val	Asp	Asp	Ile	Thr	Val	Val	Thr	His	
			180					185					190			
gtc	gcg	caa	gac	tgt	aac	cac	gcg	tct	gtt	gac	tgg	cag	gtg	gtg	gcc	624
Val	Ala	Gln	Asp	Cys	Asn	His	Ala	Ser	Val	Asp	Trp	Gln	Val	Val	Ala	
		195					200					205				
aat	ggt	gat	gtc	agc	gtt	gaa	ctg	cgt	gat	gcg	gat	caa	cag	gtg	gtt	672
Asn	Gly	Asp	Val	Ser	Val	Glu	Leu	Arg	Asp	Ala	Asp	Gln	Gln	Val	Val	
	210					215					220					
gca	act	gga	caa	ggc	act	agc	ggg	act	ttg	caa	gtg	gtg	aat	ccg	cac	720
Ala	Thr	Gly	Gln	Gly	Thr	Ser	Gly	Thr	Leu	Gln	Val	Val	Asn	Pro	His	
	225				230					235					240	
ctc	tgg	caa	ccg	ggt	gaa	ggt	tat	ctc	tat	gaa	ctg	tgc	gtc	aca	gcc	768
Leu	Trp	Gln	Pro	Gly	Glu	Gly	Tyr	Leu	Tyr	Glu	Leu	Cys	Val	Thr	Ala	
				245					250					255		
aaa	agc	cag	aca	gag	tgt	gat	atc	tac	ccg	ctt	cgc	gtc	ggc	atc	cgg	816
Lys	Ser	Gln	Thr	Glu	Cys	Asp	Ile	Tyr	Pro	Leu	Arg	Val	Gly	Ile	Arg	
			260					265					270			
tca	gtg	gca	gtg	aag	ggc	gaa	cag	ttc	ctg	att	aac	cac	aaa	ccg	ttc	864
Ser	Val	Ala	Val	Lys	Gly	Glu	Gln	Phe	Leu	Ile	Asn	His	Lys	Pro	Phe	
		275					280					285				
tac	ttt	act	ggc	ttt	ggt	cgt	cat	gaa	gat	gcg	gac	ttg	cgt	ggc	aaa	912
Tyr	Phe	Thr	Gly	Phe	Gly	Arg	His	Glu	Asp	Ala	Asp	Leu	Arg	Gly	Lys	
	290					295					300					
gga	ttc	gat	aac	gtg	ctg	atg	gtg	cac	gac	cac	gca	tta	atg	gac	tgg	960
Gly	Phe	Asp	Asn	Val	Leu	Met	Val	His	Asp	His	Ala	Leu	Met	Asp	Trp	
	305				310					315					320	
att	ggg	gcc	aac	tcc	tac	cgt	acc	tcg	cat	tac	cct	tac	gct	gaa	gag	1008
Ile	Gly	Ala	Asn	Ser	Tyr	Arg	Thr	Ser	His	Tyr	Pro	Tyr	Ala	Glu	Glu	
			325						330					335		
atg	ctc	gac	tgg	gca	gat	gaa	cat	ggc	atc	gtg	gtg	att	gat	gaa	act	1056
Met	Leu	Asp	Trp	Ala	Asp	Glu	His	Gly	Ile	Val	Val	Ile	Asp	Glu	Thr	
			340					345					350			
gct	gct	gtc	ggc	ttt	aac	ctc	tct	tta	ggc	att	ggt	ttc	gaa	gcg	ggc	1104
Ala	Ala	Val	Gly	Phe	Asn	Leu	Leu	Gly	Ile	Gly		Phe	Glu	Ala	Gly	
		355					360					365				
aac	aag	ccg	aaa	gaa	ctg	tac	agc	gaa	gag	gca	gtc	aac	ggg	gaa	act	1152
Asn	Lys	Pro	Lys	Glu	Leu	Tyr	Ser	Glu	Glu	Ala	Val	Asn	Gly	Glu	Thr	
	370					375					380					
cag	caa	gcg	cac	tta	cag	gcg	att	aaa	gag	ctg	ata	gcg	cgt	gac	aaa	1200
Gln	Gln	Ala	His	Leu	Gln	Ala	Ile	Lys	Glu	Leu	Ile	Ala	Arg	Asp	Lys	
					390					395					400	
aac	cac	cca	agc	gtg	gtg	atg	tgg	agt	att	gcc	aac	gaa	ccg	gat	acc	1248
Asn	His	Pro	Ser	Val	Val	Met	Trp	Ser	Ile	Ala	Asn	Glu	Pro	Asp	Thr	
				405					410					415		
cgt	ccg	caa	ggt	gca	cgg	gaa	tat	ttc	gcg	cca	ctg	gcg	gaa	gca	acg	1296
Arg	Pro	Gln	Gly	Ala	Arg	Glu	Tyr	Phe	Ala	Pro	Leu	Ala	Glu	Ala	Thr	
			420					425					430			
cgt	aaa	ctc	gac	ccg	acg	cgt	ccg	atc	acc	tgc	gtc	aat	gta	atg	ttc	1344
Arg	Lys	Leu	Asp	Pro	Thr	Arg	Pro	Ile	Thr	Cys	Val	Asn	Val	Met	Phe	
		435					440					445				

tgc	gac	gct	cac	acc	gat	acc	atc	agc	gat	ctc	ttt	gat	gtg	ctg	tgc	1392
Cys	Asp	Ala	His	Thr	Asp	Thr	Ile	Ser	Asp	Leu	Phe	Asp	Val	Leu	Cys	
	450					455					460					
ctg	aac	cgt	tat	tac	gga	tgg	tat	gtc	caa	agc	ggc	gat	ttg	gaa	acg	1440
Leu	Asn	Arg	Tyr	Tyr	Gly	Trp	Tyr	Val	Gln	Ser	Gly	Asp	Leu	Glu	Thr	
	465				470					475					480	
gca	gag	aag	gta	ctg	gaa	aaa	gaa	ctt	ctg	gcc	tgg	cag	gag	aaa	ctg	1488
Ala	Glu	Lys	Val	Leu	Glu	Lys	Glu	Leu	Leu	Ala	Trp	Gln	Glu	Lys	Leu	
				485					490					495		
cat	cag	ccg	att	atc	atc	acc	gaa	tac	ggc	gtg	gat	acg	tta	gcc	ggg	1536
His	Gln	Pro	Ile	Ile	Ile	Thr	Glu	Tyr	Gly	Val	Asp	Thr	Leu	Ala	Gly	
			500					505					510			
ctg	cac	tca	atg	tac	acc	gac	atg	tgg	agt	gaa	gag	tat	cag	tgt	gca	1584
Leu	His	Ser	Met	Tyr	Thr	Asp	Met	Trp	Ser	Glu	Glu	Tyr	Gln	Cys	Ala	
		515					520					525				
tgg	ctg	gat	atg	tat	cac	cgc	gtc	ttt	gat	cgc	gtc	agc	gcc	gtc	gtc	1632
Trp	Leu	Asp	Met	Tyr	His	Arg	Val	Phe	Asp	Arg	Val	Ser	Ala	Val	Val	
	530					535					540					
ggg	gaa	cag	gta	tgg	aat	ttc	gcc	gat	ttt	gcg	acc	tcg	caa	ggc	ata	1680
Gly	Glu	Gln	Val	Trp	Asn	Phe	Ala	Asp	Phe	Ala	Thr	Ser	Gln	Gly	Ile	
	545				550					555					560	
ttg	cgc	gtt	ggc	ggg	aac	aag	aaa	ggg	atc	ttc	act	cgc	gac	cgc	aaa	1728
Leu	Arg	Val	Gly	Gly	Asn	Lys	Lys	Gly	Ile	Phe	Thr	Arg	Asp	Arg	Lys	
				565					570					575		
ccg	aag	tcg	gcg	gct	ttt	ctg	ctg	caa	aaa	cgc	tgg	act	ggc	atg	aac	1776
Pro	Lys	Ser	Ala	Ala	Phe	Leu	Leu	Gln	Lys	Arg	Trp	Thr	Gly	Met	Asn	
			580					585					590			
ttc	ggg	gaa	aaa	ccg	cag	cag	gga	ggc	aaa	caa	tga					1812
Phe	Gly	Glu	Lys	Pro	Gln	Gln	Gly	Gly	Lys	Gln	*					
		595					600									

<210> 106
 <211> 603
 <212> PRT
 <213> Escherichia coli

<300>
 <308> Genbank #S69414
 <309> 1994-09-23

<400>	106															
Met	Leu	Arg	Pro	Val	Glu	Thr	Pro	Thr	Arg	Glu	Ile	Lys	Lys	Leu	Asp	
1				5					10					15		
Gly	Leu	Trp	Ala	Phe	Ser	Leu	Asp	Arg	Glu	Asn	Cys	Gly	Ile	Asp	Gln	
			20					25					30			
Arg	Trp	Trp	Glu	Ser	Ala	Leu	Gln	Glu	Ser	Arg	Ala	Ile	Ala	Val	Pro	
		35					40					45				
Gly	Ser	Phe	Asn	Asp	Gln	Phe	Ala	Asp	Ala	Asp	Ile	Arg	Asn	Tyr	Ala	
	50				55						60					
Gly	Asn	Val	Trp	Tyr	Gln	Arg	Glu	Val	Phe	Ile	Pro	Lys	Gly	Trp	Ala	
	65				70				75						80	
Gly	Gln	Arg	Ile	Val	Leu	Arg	Phe	Asp	Ala	Val	Thr	His	Tyr	Gly	Lys	
			85						90					95		
Val	Trp	Val	Asn	Asn	Gln	Glu	Val	Met	Glu	His	Gln	Gly	Gly	Tyr	Thr	
			100					105					110			
Pro	Phe	Glu	Ala	Asp	Val	Thr	Pro	Tyr	Val	Ile	Ala	Gly	Lys	Ser	Val	
			115				120					125				

Arg	Ile	Thr	Val	Cys	Val	Asn	Asn	Glu	Leu	Asn	Trp	Gln	Thr	Ile	Pro
130						135					140				
Pro	Gly	Met	Val	Ile	Thr	Asp	Glu	Asn	Gly	Lys	Lys	Lys	Gln	Ser	Tyr
145						150				155					160
Phe	His	Asp	Phe	Phe	Asn	Tyr	Ala	Gly	Ile	His	Arg	Ser	Val	Met	Leu
				165					170					175	
Tyr	Thr	Thr	Pro	Asn	Thr	Trp	Val	Asp	Asp	Ile	Thr	Val	Val	Thr	His
			180					185					190		
Val	Ala	Gln	Asp	Cys	Asn	His	Ala	Ser	Val	Asp	Trp	Gln	Val	Val	Ala
		195					200					205			
Asn	Gly	Asp	Val	Ser	Val	Glu	Leu	Arg	Asp	Ala	Asp	Gln	Gln	Val	Val
	210					215					220				
Ala	Thr	Gly	Gln	Gly	Thr	Ser	Gly	Thr	Leu	Gln	Val	Val	Asn	Pro	His
225						230				235					240
Leu	Trp	Gln	Pro	Gly	Glu	Gly	Tyr	Leu	Tyr	Glu	Leu	Cys	Val	Thr	Ala
				245					250					255	
Lys	Ser	Gln	Thr	Glu	Cys	Asp	Ile	Tyr	Pro	Leu	Arg	Val	Gly	Ile	Arg
			260					265					270		
Ser	Val	Ala	Val	Lys	Gly	Glu	Gln	Phe	Leu	Ile	Asn	His	Lys	Pro	Phe
		275					280					285			
Tyr	Phe	Thr	Gly	Phe	Gly	Arg	His	Glu	Asp	Ala	Asp	Leu	Arg	Gly	Lys
	290					295					300				
Gly	Phe	Asp	Asn	Val	Leu	Met	Val	His	Asp	His	Ala	Leu	Met	Asp	Trp
305					310					315					320
Ile	Gly	Ala	Asn	Ser	Tyr	Arg	Thr	Ser	His	Tyr	Pro	Tyr	Ala	Glu	Glu
				325					330					335	
Met	Leu	Asp	Trp	Ala	Asp	Glu	His	Gly	Ile	Val	Val	Ile	Asp	Glu	Thr
			340					345					350		
Ala	Ala	Val	Gly	Phe	Asn	Leu	Ser	Leu	Gly	Ile	Gly	Phe	Glu	Ala	Gly
		355					360					365			
Asn	Lys	Pro	Lys	Glu	Leu	Tyr	Ser	Glu	Glu	Ala	Val	Asn	Gly	Glu	Thr
	370					375					380				
Gln	Gln	Ala	His	Leu	Gln	Ala	Ile	Lys	Glu	Leu	Ile	Ala	Arg	Asp	Lys
385					390					395					400
Asn	His	Pro	Ser	Val	Val	Met	Trp	Ser	Ile	Ala	Asn	Glu	Pro	Asp	Thr
				405					410					415	
Arg	Pro	Gln	Gly	Ala	Arg	Glu	Tyr	Phe	Ala	Pro	Leu	Ala	Glu	Ala	Thr
			420					425					430		
Arg	Lys	Leu	Asp	Pro	Thr	Arg	Pro	Ile	Thr	Cys	Val	Asn	Val	Met	Phe
		435					440					445			
Cys	Asp	Ala	His	Thr	Asp	Thr	Ile	Ser	Asp	Leu	Phe	Asp	Val	Leu	Cys
	450					455					460				
Leu	Asn	Arg	Tyr	Tyr	Gly	Trp	Tyr	Val	Gln	Ser	Gly	Asp	Leu	Glu	Thr
465					470					475					480
Ala	Glu	Lys	Val	Leu	Glu	Lys	Glu	Leu	Leu	Ala	Trp	Gln	Glu	Lys	Leu
				485					490					495	
His	Gln	Pro	Ile	Ile	Ile	Thr	Glu	Tyr	Gly	Val	Asp	Thr	Leu	Ala	Gly
		500						505					510		
Leu	His	Ser	Met	Tyr	Thr	Asp	Met	Trp	Ser	Glu	Glu	Tyr	Gln	Cys	Ala
		515					520					525			
Trp	Leu	Asp	Met	Tyr	His	Arg	Val	Phe	Asp	Arg	Val	Ser	Ala	Val	Val
	530					535					540				
Gly	Glu	Gln	Val	Trp	Asn	Phe	Ala	Asp	Phe	Ala	Thr	Ser	Gln	Gly	Ile
545					550					555					560
Leu	Arg	Val	Gly	Gly	Asn	Lys	Lys	Gly	Ile	Phe	Thr	Arg	Asp	Arg	Lys
				565					570					575	
Pro	Lys	Ser	Ala	Ala	Phe	Leu	Leu	Gln	Lys	Arg	Trp	Thr	Gly	Met	Asn
			580					585					590		
Phe	Gly	Glu	Lys	Pro	Gln	Gln	Gly	Gly	Lys	Gln					
		595					600								

<210> 107

<211> 277

<212> DNA

<213> Artificial Sequence

<220>
<223> Nopaline Synthase Terminator Sequence

<300>
<308> U09365
<309> 1995-10-17

<400> 107
gagctcgaat tcccccgatc gttcaaacat ttggcaataa agtttcttaa gattgaatcc 60
tggtgcccgt cttgcgatga ttatcatata atttctgttg aattacgtta agcatgtaat 120
aattaacatg taatgcatga cgttatttat gagatgggtt tttatgatta gaggccccga 180
attatacatt taatagcgga tagaaaacaa aatatagcgc gcaaaactagg ataaattatc 240
gcgcgcgggt tcactctatgt tactagatcg ggaattc 277

<210> 108
<211> 3451
<212> DNA
<213> Artificial Sequence

<220>
<223> HindIII Fragment containing the beta-glucuronidase coding sequence, the rDNA intergenic spacer, and the Mast1 sequence

<400> 108
aagcttgacc tggaatatcg cgagtaaact gaaaatcacg gaaaatgaga aatacacact 60
ttaggacgtg aaatatggcg agggaaaactg aaaaagggtgg aaaatttaga aatgtccact 120
gtaggacgtg gaatatggca agaaaactga aaatcatgga aaatgagaaa catccacttg 180
acgacttgaa aaatgacgaa atcactaaaa aacgtgaaaa atgagaaaatg cacactgaag 240
gactccgcgg gaattcgatt gtgctagcca atgtttaaca agatgtcaag cacaatgaat 300
gttggtgggt ggtggtcgtg gctggcgggt gtggaaaatt gcgggtggtc gagcggtagt 360
gatcggcgat ggttgggtgt tgcagcgggt tttgatatcg gaatcactta tgggtggtgt 420
cacaatggag gtgcgtcatg gttattgggt gttggctatc tatatatatt tataataata 480
ttaagtattt tacctatttt ttacatattt tttattaaat ttatgcattg tttgtatttt 540
taaagtattt ttatcgtact tgttttataa aatattttat tattttatgt gttatattat 600
tacttgatgt attggaaatt ttctccattg tttttcttat atttataata attttcttat 660
ttttttttgt tttattatgt attttttcgt tttataataa atatttatta aaaaaaatat 720
tattttttgt aaatatatca tttacaatgt ttaaaagtca tttgtgaata tattagctaa 780
gttgtacttc ttttgggttg tttgggtgct tacatgtcta ttatgattct ctggccaaaa 840
catgtctact cctgtcactt ggggtttttt ttttaagaca taatcactag tgattatatc 900
tagactgaag gcgggaaacg acaatctgat catgagcggg gaattaaggg agtcacgtta 960
tgaccccgcg cgatgacgcg ggacaagcgg ggaactgaca gaacgcgaac 1020
gttgaaggag ccactcagcc gcgggtttct ggagttaaat gagctaagca catacgtcag 1080
aaaccattat tgcgcgttca aaagtcgcct aaggtcacta tcagctagca aatatttctt 1140
gtcaaaaaatg ctccactgac gttccataaa tttccctcgg tatccaatta gagtctcata 1200
ttcactctca atccaaataa tctgcaccgg atctcgagat cgaattcccg cggccgcgaa 1260
ttcactagtg gatccccggg tacggctcagt cccttatgtt acgtcctgta gaaaacccaa 1320
cccgtgaaat caaaaaaactc gacggcctgt gggcattcag tctggatcgc gaaaactgtg 1380
gaattgagca gcgttggtgg gaaagcgcgt tacaagaaag ccgggcaatt gctgtgccag 1440
gcagttttaa cgatcagttc gccgatgcag atatttcgta ttatgtgggc aacgtctggg 1500
atcagcgca agtctttata ccgaaagggt gggcaggcca gcgtatcgt ctgctttctg 1560
atgcggtcac tcattacggc aaagtgtggg tcaataatca ggaagtgatg gagcatcagg 1620
gcggctatac gccatttgaa gccgatgtca cgccgtatgt tattgccggg aaaagtgtac 1680
gtatcacagt ttgtgtgaac aacgaactga actggcagac tatcccgccg ggaatgggtg 1740
ttaccgacga aaacggcaag aaaaagcagt cttacttcca tgatttcttt aactacgccg 1800
ggatccatcg cagcgtaatg ctctacacca cgccgaacac ctgggtggac gatataccg 1860
tggtgacgca tgtcgcgcaa gactgtaacc acgcgtctgt tgactggcag gtggtggcca 1920
atggtgatgt cagcgttgaa ctgctgatg cggatcaaca ggtggttgca actggacaag 1980
gcaccagcgg gactttgcaa gtggtgaatc cgcacctctg gcaaccgggt gaaggttatc 2040
tctatgaact gtacgtcaca gccaaaagcc agacagagt tgatatctac ccgctgcgcg 2100
tcggcatccg gtcagtggca gtgaaggcg aacagttcct gatcaaccac aaacgcgtct 2160
actttactgg ctttggccgt catgaagatg cggatttgcg cggaagga ttcgataacg 2220
tgctgatggt gcacgatcac gcattaatgg actggattgg ggccaactcc taccgtacct 2280
cgattacccc ttacgctgaa gagatgctcg actgggcaga tgaacatggc atcggtggtg 2340
ttgatgaaac tgcagctgtc ggctttaacc tctcttttag cattgggttc gaagcgggca 2400
acaagccgaa agaactgtac agcgaagagg cagtcaacgg ggaaactcag caggcgcact 2460
tacaggcgat taaagagctg atagcgcgtg acaaaaaacca ccaagcgtg gtgatgtgga 2520

gtattgccaa	cgaaccggat	acccgtccgc	aaggtgcacg	ggaatatttc	gcgccactgg	2580
cggaagcaac	gcgtaaaactc	gatccgacgc	gtccgatcac	ctgcgtcaat	gtaatgttct	2640
gcgacgctca	caccgataacc	atcagcgatc	tctttgatgt	gctgtgcctg	aaccgttatt	2700
acggttggtg	tgtccaaagc	ggcgatttgg	aaacggcaga	gaaggtactg	gaaaaagaac	2760
ttctggcctg	gcaggagaaa	ctgcatcagc	cgattatcat	caccgaatac	ggcgtggata	2820
cgtagccggg	gctgcactca	atgtacaccg	acatgtggag	tgaagagtat	cagtgtgcat	2880
ggctggatat	gtatcaccgc	gtcttttgatc	gcgtcagcgc	cgctcgtcgg	gaacagggtat	2940
ggaatttcgc	cgattttgcg	acctcgcaag	gcatattgcg	cgttggcggg	aacaagaagg	3000
ggatcttcac	ccgcgaccgc	aaaccgaagt	cggcggcctt	tctgctgcaa	aaacgctgga	3060
ctggcatgaa	cttcgggtgaa	aaaccgcagc	agggaggcaa	acaatgaatc	aacaactctc	3120
ctggcgaccc	atcgctgggt	acagcctcgg	gaattgctga	ccgagctcga	atttccccga	3180
tcgttcaaac	atttggaat	aaagtttctt	aagattgaat	cctgttgccg	gtcttgcat	3240
gattatcata	taattttctgt	tgaattacgt	taagcatgta	ataattaaca	tgtaatgcac	3300
gacgttattt	tttttatgat	tagagtcccg	ggataaatta	caattataca	tttaatacgc	3360
gatagaaaac	aaaatatagc	gcgcaaaacta	t	tcgcgcgcgg	tgatcatctat	3420
gttactagat	cggaatttcg	atatcaagct				3451

<210> 109
 <211> 14627
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pAg11a Plasmid

<400> 109						
catgccaaacc	acagggttcc	cctcggggtc	aaagtacttt	gateccaaccc	ctccgctgct	60
atagtgcagt	cggtcttctga	cgttcagtg	agccgtcttc	tgaaaacgcac	atgtcgcaca	120
agtcctaagt	tacgcgacag	gctgccgccc	tgcccttttc	ctggcggttt	cttgtcgcgt	180
gttttagtgc	cataaagtag	aatacttgcc	actagaaccg	gagacattac	gccatgaaca	240
agagcgccgc	cgctggcctg	ctgggctatg	cccgcgtcag	caccgacgac	caggacttga	300
ccaaccaacg	ggccgaactg	cacgcggccg	gctgcaccaa	gctgttttcc	gagaagatca	360
ccggcaccag	gcgcgacccg	ccggagctgg	ccaggatgct	tgaccaccta	cgccctggcg	420
acgttgtgac	agtgaccagg	ctagaccgcc	tgcccgccag	caccgcgcac	ctactggaca	480
ttgccgagcg	catccaggag	gcccggcgccg	gcctgcgtag	cctggcagag	ccgtggggccg	540
acaccaccac	gcgggcgggc	cgcatggtgt	tgaccgtgtt	cgccggcatt	gccgagttcg	600
agcgttccct	aatcatcgac	cgcacccgga	gcgggcgcga	ggccgcgaag	gcccaggccg	660
tgaagtttgg	ccccgcctt	acctcacc	cggcacagat	cgcgacgcgc	cgcgagctga	720
tcgaccagga	agggccgacc	gtgaaagagg	cggctgcact	gcttggcggtg	catcgctcga	780
ccctgtaccg	cgcacttgag	cgcagcgagg	aagtgcgcgc	caccgaggcc	aggcggcgcg	840
gtgccttccg	tgaggacgca	ttgaccgagg	ccgacgccc	ggcggccgcc	gagaatgaac	900
gccaagagga	acaagcatga	aaccgcacca	ggacggccag	gacgaaccgt	ttttcattac	960
cgaagagatc	gaggcggaga	tgatcgccgc	cgggtacgtg	ttcgagccgc	ccgcgcacgt	1020
ctcaaccgtg	cggtctgcag	aaatcctggc	cggtttgtct	gatgccaaagc	tgggcgccctg	1080
gcccgcagc	ttggccgctg	aagaaaccga	gcgcgcgcgt	ctaaaaagggt	gatgtgtatt	1140
tgagtataaac	agcttgccgc	atgcggtcgc	tgcttatatg	atgcgatgag	tataataaca	1200
aatacgcaag	gggaacgcac	gaaggttatc	gctgtactta	accagaaagg	cggttcaggc	1260
aagacgacca	tcgcaaccca	tctagccccc	gcccgtgcaac	tcgcccggggc	cgatgttctg	1320
ttagtccgatt	ccgatcccca	gggcagtgcc	cgcgattggg	cggccgtgcg	ggaagatcaa	1380
ccgctaaccg	ttgtcgccat	cgaccgcccc	acgattgacc	gcgacgtgaa	ggccatcggc	1440
cggcgcgact	tcgtagtgat	cgacggagcg	ccccaggccg	cggacttggc	tgtgtcccg	1500
atcaaggcag	ccgacttcgt	gctgattccg	gtgcagccaa	gcccctacga	catatggg	1560
accgcgcacc	tggtggagct	gggttaagcag	cgcattgagg	tcacggatgg	aaggctacaa	1620
gcggcctttg	tcgtgtccgc	ggcgatcaaa	ggcacgcgca	tcggcggtga	ggttgccgag	1680
gcgctggccg	ggtacgagct	gcccattctt	gagtcgccga	tcacgcagcg	cgtgagctac	1740
ccaggcactg	ccgcgcgcgg	cacaaccgtt	cttgaatcag	aaccgcagg	cgacgctgcc	1800
cgcgaggtcc	agggcgtggc	cgtgaaaatt	aaatcaaaac	tcatttgagt	taatgaggt	1860
aagagaaaaat	gagcaaaagc	acaaacacgc	taagtgcggg	ccgtccgagc	gcacgcagca	1920
gcaaggtctgc	aacgttggcc	agcctggcag	acacgccagc	catgaagcgg	gtcaactttc	1980
agttgcgcgc	ggaggtcac	accaagctga	agatgtacgc	ggtacgcaa	ggcaagacca	2040
ttaccgagct	gctatctgaa	tacatcgccg	agctaccaga	gtaaatgagc	aaatgaataa	2100
atgagttagt	gaatttttagc	ggctaaagga	ggcgccatgg	aaaatcaaga	acaaccaggc	2160
accgacgcgc	tggaatccc	catgtgtgga	ggacggggcg	gttggccagg	cgtaagcgc	2220
tggtgtgtct	gcccgccttg	caatggcact	ggaaaccccc	agcccaggga	atcggcgtga	2280
cggtcgcaaa	ccatccggcc	cggtacaaat	cggcgccggcg	ctgggtgatg	acctggtgga	2340
gaagttagag	ccgcgcagg	ccgcccagcg	gcaacgcac	gaggcagaag	cacgcccgcg	2400
tgaatcgtgg	caagcggccg	ctgatcgaa	ccgcaaaagaa	tcccggcaac	cgccggcagc	2460

cggtgcgcgc	tgcattagga	agccgcccac	gggagcagag	caaccagatt	ttttcgttcc	2520
gatgctctat	gacgtgggca	cccgcgatag	tgcgagcatc	atggacgtgg	ccgttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	ggtgatccgc	tacgagcttc	cagacgggca	2640
cgtagagggt	tccgcagggc	cgcccgccat	ggccagtggt	tgggattacg	acctgggtact	2700
gatggcggtt	tcccatctaa	ccgaatccat	gaaccgatac	cggaaggga	agggagacaa	2760
gcccggccgc	gtgttccgtc	cacacgttgc	ggacgtactc	aagtctctgc	ggcgagccga	2820
tggcggaag	cagaaagacg	acctggtaga	aacctgcatt	cggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cggccgcctg	gtgacggtat	ccgaggggtga	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaaacc	gggcggccgg	agtacatcga	3000
gatcgagcta	gctgattgga	tgtaccgcga	gatcacagaa	ggcaagaacc	cggacgtgct	3060
gacggttcac	cccgttact	ttttgatcga	tcccgcctc	ggccggtttc	tctaccgcct	3120
ggcagccgc	gcccagggca	aggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
cagtgccagc	gcccggagagt	tcaagaagtt	ctgtttccacc	gtgcgcaagc	tgatcgggtc	3240
aatgacctg	ccggagtacg	atttgaagg	ggaggcgggg	caggctggcc	cgatccagt	3300
catgcgctac	cgcaacctga	tgcagggcga	agctccgcgc	ggttccctaat	gtacggagca	3360
gatgctaggg	caaattgccc	tagcagggga	aaaaggtcga	aaaggtctct	ttcctgtgga	3420
tagcacgtac	attgggaacc	caaagccgta	cattgggaac	cggaaccctg	acattgggaa	3480
cccaaagccg	tacattggga	accggtcaca	catgtaaagt	actgatataa	aagagaaaaa	3540
aggcgatttt	tccgcctaaa	actctttaaa	acttattaaa	actcttaaaa	cccgctggc	3600
ctgtgcataa	ctgtctggcc	agcgcacagc	cgaagagctg	caaaaagcgc	ctaccctcg	3660
gtegetgcgc	tccctacgce	cgcgcgcttc	cgctcgccct	atcgccggcc	ctggccgctc	3720
aaaaatggct	ggcctacggc	caggcaatct	accagggcgc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	cggcgccccc	atcaaggcac	cctgcctcgc	gcgtttcggg	gatgacgggt	3840
aaaacctctg	acacatgcag	ctcccggaga	cggtcacagc	ttgtctgtaa	cgggatgcgc	3900
ggagcagaca	agcccgctcag	ggcgcgctcag	cgggtgttgg	cgggtgtcgg	ggcgagccca	3960
tgaccagtc	acgtagcgat	agcggagtg	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgcacc	atatgcggtg	tgaataaccg	cacagatgcg	taaggagaaa	4080
ataccgcctc	aggcgctctt	ccgcttccct	gctcactgac	tcgctgcgct	cggtcgttcg	4140
gctgcggcga	gcggtatcag	ctcactcaaa	ggcggtataa	cggttatcca	cagaatcagg	4200
ggataacgca	tgtagagaaa	tgtagagaaa	aggccagcaa	aaggccagga	accgtaaaaa	4260
ggccgcggtg	ctggcggttt	tccataggct	ccgccccct	gacgagcatc	acaaaaatcg	4320
acgctcaagt	cagaggtggc	gaaaccgcag	aggactataa	agataccagg	cgtttcccc	4380
tggaaagctc	ctcgtgcgct	ctcctgttcc	gacctgcgc	cttaccggat	acctgtccgc	4440
ctttctccct	tccgggaagcg	tggcgctttc	tcatagctca	cgctgtagg	atctcagttc	4500
gggtgtaggc	gttcgctcca	agctgggctg	tgtgcacgaa	cccccggtc	agcccgaccg	4560
ctgcgcctta	tccggtaact	atcgctctga	gtccaaccgc	gtaagacacg	acttatcgcc	4620
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	4680
gttcttgaag	tgggtggccta	actacggcta	cactagaagg	acagtatttg	gtatctgcgc	4740
tctgctgaag	ccagttacct	tccgaaaaag	agttggtagc	tcttgatocg	gcaaaacaaa	4800
caccgctggg	agcgggtgggt	tttttgtttg	caagcagcag	attacgcgca	gaaaaaaagg	4860
atctcaagaa	gatcctttga	tctttctctc	ggggtctgac	gctcagtgga	acgaaaaact	4920
acgttaaggg	atcttgggtc	tgcattctag	gctactaaa	aattcatcca	gtaaaaata	4980
atattttatt	ttctcccaat	caggcttgat	ccccagtaag	tcaaaaaata	gctcgacata	5040
ctgttcttcc	ccgatattct	ccctgatcga	ccggacgcag	aaggcaatgt	cataccactt	5100
gtccgcctcg	ccgcttctcc	caagatcaat	aaagccactt	actttgccat	ctttcacaaa	5160
gatgttgctg	tctcccaggt	cgccgtggga	aaagacaagt	tctcttccgg	gcttttccgt	5220
ctttaaaaaa	tcatcacgct	cgccgggatc	tttaaatgga	gtgtcttctt	cccagttttc	5280
gcaatccaca	tccggccagat	cgttattcag	taagtaatcc	aattcggcta	agcggctgtc	5340
taagctattc	gtatagggac	aatccgatat	gtcgatggag	tgaagagacc	tgatgcactc	5400
cgcatacagc	togataatct	tttcagggct	ttgttcatct	tcatactctt	ccgagcaaa	5460
gacgccatcg	gocctcactc	tgagcagatt	gctccagcca	tcatgcccgt	caaagtgcag	5520
gacctttgga	acaggcagct	ttccttccag	ccatagcatc	atgtcctttt	cccgttccac	5580
atcatagggt	gtccctttat	accggctgtc	cgctcatttt	aaatatagg	tttcattttc	5640
tcccaccagc	ttatatacct	tagcaggaga	cattccctcc	gtatctttta	cgcagcggtg	5700
tttttcgata	agttttttca	attccgggtg	tattctcatt	ttagccattt	attattttct	5760
tctctttttc	tacagttatt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
aattcactgt	tccctgcatt	ctaaaacctt	aaataccaga	aaacagcttt	ttcaagttg	5880
ttttcaaagt	tggcggtata	catagtatcg	acggagccga	ttttgaaacc	gcggtgatca	5940
caggcgcaaa	cgctctgtca	togttacaat	caacatgcta	ccctccgcga	gatcatccgt	6000
gtttcaaaac	cggcagctta	gttgccgctt	ttccgaatag	catcggtaac	atgagcaaac	6060
tctgcgcgct	tacaacggct	ctcccgctga	cgcgctcccg	gactgatggg	ctgcctgtat	6120
cgagtggtga	ttttgtgccc	agctgcccgt	cggggagctg	ttggctggct	gggtggcagga	6180
tatatgttgg	tgtaaacaaa	ttgacgctta	gacaacttaa	taacacattg	cggaggtttt	6240
taattgtatg	gaattaacgc	gaattaatcc	gggggatctg	gatttttagta	ctggattttg	6300
gtttttaggaa	ttagaaattt	tattgataga	agttattttac	aaatacaaat	acataactaag	6360
ggtttctttat	atgctcaaca	catgagcgaa	acctataagg	aaccctaatt	cccttagctg	6420
ggaactactc	acacattatt	atggagaaac	tcgagtcaaa	tctcggtgac	gggcaggaac	6480

ggacggggcg	gtaccggcag	gctgaagtcc	agctgccaga	aacccacgtc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgccctcgtc	6600
atgcgacgc	tcgggtcggt	gggcagcccg	atgacagcga	ccacgctctt	gaagccctgt	6660
gcctccagg	acttcagcag	gtgggtgtag	agcgtggagc	ccagtcccgt	ccgctgggtg	6720
cggggggaga	cgtacacggg	cgactcggcc	gtccagtcgt	aggcgttgcg	tgccctccag	6780
gggcccgcgt	aggcgatgcc	ggcgacctcg	ccgtccacct	cggcgacgag	ccagggatag	6840
cgctcccgc	gacggacgag	gtcgtccgtc	cactcctgcg	gttcctgctg	ctcggtagcg	6900
aagttgacgg	tgcttgtctc	gatgtagtgg	ttgacgatgg	tgacagaccg	cgccatgtcc	6960
gcctcgggtg	cacggcggat	gtcggccggg	cgtcgttctg	ggctcatggg	agactcgaga	7020
gagatagatt	tgtagagaga	gactgggtgat	ttcagcgtgt	cctctccaaa	tgaaatgaac	7080
ttccttatat	agaggaaggt	cttgcgaagg	atagtgggat	tgtgcgtcat	cccttacgtc	7140
agtggagata	tcacatcaat	ccacttgctt	tgaagacgtg	gttggaacgt	cttctttttc	7200
cacgatgtct	ctcgtgggtg	gggggtccatc	tttgggacca	ctgtcggcag	aggcatcttg	7260
aacgatagcc	tttcccttat	cgcaatgatg	gcatttgtag	gtgccacctt	ccttttcttg	7320
tgctcttttg	atgaagtggc	agatagctgg	gcaattggaat	ccgaggaggt	ttcccgatat	7380
taccctttgt	tgaaaagtct	caatagccct	ttgggtcttct	gagactgtat	ctttgatatt	7440
cttgaggtag	acgagagtgt	cgtgctccac	catgttatca	catcaatcca	cttgctttga	7500
agacgtgggt	ggaaacgtct	ctttttccac	gatgctcctc	gtgggtgggg	gtccatcttt	7560
gggaccactg	tcggcagagg	catcttgaac	gatagccttt	ccttttatcg	aatgatggca	7620
tttgtaggtg	ccaccttctc	tttctactgt	ccttttgatg	aagtacaga	tagctgggca	7680
atggaatccg	aggaggtttc	ccgatattac	cctttgttga	aaagtctcaa	tagccctttg	7740
gtcttctgag	actgtatctt	tgatattctt	ggagtagacg	agagtgtcgt	gctccaccat	7800
gttggcaagc	tgctctagcc	aatacgcaaa	ccgcctctcc	ccgcgcgttg	gccgattcat	7860
taatgcagct	ggcacgacag	gtttcccgcg	tggaaagcgg	gcagtgagcg	caacgcaatt	7920
aatgtgagtt	agctcactca	ttaggcaccc	caggctttac	actttatgct	tcgggctcgt	7980
atgttgtgtg	gaatttgtgag	cggataacaa	tttcacacag	gaaacagcta	tgaccatgat	8040
tacgaattcg	agccttgact	agagggtcga	cggtatacac	acatgataag	atacattgat	8100
gagtttgagc	aaaccacaac	tagaatgcag	tgaaaaaaat	gctttatttg	tgaaatttgt	8160
gatgctattg	ctttatattgt	aaccattata	agctgcaata	aacaagtggg	ggtgggcgaa	8220
gaactccagc	atgagatccc	cgcgctggag	gatcatccag	ccggcgtccc	ggaaaacgat	8280
tcggaagccc	aacctttcat	agaaggcggc	ggtggaatcg	aaatctcgta	gcacgtgtca	8340
gtcctgtctc	tcggccacga	agtgcacgca	gttgccggcc	gggtcgcgca	gggcgaactc	8400
ccgccccac	ggctgtcgcg	cgatctcggt	catggccggc	ccggaggcgt	cccggaagtt	8460
cgtggacacg	acctccgacc	actcggcgta	cagctcgtcc	aggccgcgca	cccacaccca	8520
ggccagggtg	ttgtccggca	ccacctgggtc	ctggacccgc	ctgatgaaca	gggtcacgtc	8580
gtcccggacc	acaccggcga	agtctgtcct	cacgagttcc	cgggagaacc	cgagccggtc	8640
ggtccagaac	tcgaccgctc	cggcgacgtc	gcgcgcggtg	agcaccggaa	cggcactggg	8700
caacttggcc	atggatccag	atttcgctca	agttagtata	aaaaagcagg	cttcaatcct	8760
gcaggaattc	gatcgacact	ctcgtctact	ccaagaatat	caaagataca	gtctcagaag	8820
accaaaaggg	tattgagact	tttcaacaaa	gggtaatatc	gggaaacctc	ctcggattcc	8880
attgcccagc	tatctgtcac	ttcatcaaaa	ggacagtaga	aaagggaagg	ggcacctaca	8940
aatgccatca	ttgcgataaa	ggaaaggcta	tcgttcaaga	tgctctgccc	gacagtgttc	9000
ccaaagattg	accccacccc	acgaggagca	tcgtggaaaa	agaagacgtt	ccaaccacgt	9060
cttcaaagca	agtggattga	tgtgataaca	tggtggagca	cgacactctc	gtctactcca	9120
agaatatcaa	agatacagtc	tcagaagacc	aaagggtctat	tgagactttt	caacaaaagg	9180
taatatcggg	aaacctcctc	ggattccatt	gcccagctat	ctgtcacttc	atcaaaagga	9240
cagtagaaaa	ggaagggtgg	acctacaaat	gccatcattg	cgataaaagg	aaggctatcg	9300
ttcaagatgc	ctctgccgac	agtggtecca	aagatggacc	cccacccacg	aggagcatcg	9360
tggaaaaaaga	agacgttcca	accacgtctt	caaagcaagt	ggattgatgt	gatattctcca	9420
ctgacgtaag	ggatgacgca	caatcccact	atccttcgca	agaccttcc	ctatataaag	9480
aagttcattt	catttgagga	ggacacgctg	aatcaccag	tctctctcta	caaattctat	9540
tctctcgagc	tttcgcagat	ccgggggggg	aatgagatat	gaaaaagcct	gaactcaccg	9600
cgacgtctgt	cgagaagttt	ctgatcgaaa	agttcgacag	cgtctccgac	ctgatgcagc	9660
tctcggaggg	cgaagaatct	cgtgcttttc	gcttcgatgt	aggaggcgtg	ggatattgtc	9720
tgccgggtaaa	tagctgcgcc	gatggtttct	acaaagatcg	ttatgtttat	cggcactttg	9780
catcgccgcg	gctcccgat	ccggaagtgc	ttgacattgg	ggagtttagc	gagagcctga	9840
cctattgcat	ctcccgcgtg	gcacaggggt	tcacgttgca	agacctgcct	gaaaccgaac	9900
tgcccgctgt	tctacaaccc	gtcgcggagg	ctatggatgc	gatcgctgcg	gccgatctta	9960
gccagacgag	cgggttcggc	ccattcggac	cgcaagggaat	cggtcaatac	actacatggc	10020
gtgatttcat	atgcgcgatt	gctgatcccc	atgtgtatca	ctggcaaaact	gtgatggacg	10080
acaccgtcag	tgcgctccgt	gcgcaggctc	tcgatgagct	gatgcttttg	gccgaggact	10140
gccccgaagt	ccggcacctc	gtgcacgcgg	atttcggctc	caacaatgtc	ctgacggaca	10200
atggccgcgt	aacagcgggt	attgactgga	gcgaggcgat	gttcggggat	tcccaatacg	10260
aggtcgccaa	catcttcttc	tggaggcgtg	ggttggcttg	tatggagcag	cagacgcgtc	10320
acttcgagcg	gaggcatccg	gagcttgca	gatcgccacg	actccgggcg	tatatgctcc	10380
gcattgggtc	tgaccaactc	tatcagagct	tggttgacgg	caatttcgat	gatgcagctt	10440
gggcgcaggg	tcgatgcgac	gcaatcgtcc	gatccggagc	cgggactgtc	gggcgtacac	10500

aaatcgcccg	cagaagcgcg	gccgtctgga	ccgatggctg	tgtagaagta	ctcgccgata	10560
gtggaaacccg	acgccccagc	actcgctccga	gggcaaagaa	atagagtaga	tgccgaccgg	10620
atctgtcgat	cgacaagctc	gagtttctcc	ataataatgt	gtgagtagtt	cccagataag	10680
ggaattaggg	ttcctatagg	gtttcgctca	tgtgttgagc	atataagaaa	cccttagtat	10740
gtatttgtat	ttgtaaaata	cttctatcaa	taaaatttct	aattcctaaa	accaaatacc	10800
agtactaaaa	tccagatccc	ccgaattaat	tccggcgtaa	ttcagatcaa	gcttgacctg	10860
gaatatcgcg	agtaaaactga	aaatcacgga	aaatgagaaa	tacacacttt	aggacgtgaa	10920
atatggcgag	gaaaactgaa	aaagggtggaa	aatttagaaa	tgtccactgt	aggacgtgga	10980
atatggcaag	aaaactgaaa	atcatggaaa	atgagaaaca	tccacttgac	gacttgaaaa	11040
atgacgaaat	cactaaaaaa	cgtgaaaaat	gagaaatgca	cactgaagga	ctccgcggga	11100
attcgattgt	gctagccaat	gtttaacaag	atgtcaagca	caatgaatgt	tggtggttgg	11160
tggtcggtgg	tggtcggtgg	ggaaaattgc	gggtggttgc	gcggtagtga	tcggcgatgg	11220
ttggtgtttg	cagcggtgtt	tgatatcgga	atcacttatg	gtggttgtca	caatggagggt	11280
gcgtcatggg	tattgggtgg	tggtcatcta	tatattttta	taataatatt	aagtatttta	11340
cttatttttt	acataatttt	tattaaattt	atgcattgtt	tgtattttta	aatagttttt	11400
atcgtaacttg	ttttataaaaa	tatttttatta	ttttatgtgt	tatatattta	cttgatgtat	11460
tggaattttt	ctccatttgt	ttttctatat	ttataataat	tttcttattt	ttttttgttaa	11520
tattatgtat	tttttcgttt	tataataaat	attttataaa	aaaaatatta	tttttgttaa	11580
atatatcatt	tacaatgttt	aaaagtcatt	tgtgaatata	ttagctaagt	tgtacttctt	11640
tttgtgcatt	tggtgttgtta	catgtctatt	atgattctct	ggccaaaaca	tgtctactcc	11700
tgtcacttgg	gttttttttt	ttaagacata	atcactagt	attatatcta	gactgaaggg	11760
gggaaacgac	aatctgatca	tgagcggaga	athtaaggag	tcacgttatg	accccgcccg	11820
atgacgcggg	acaagccgtt	ttacgtttgg	aactgacaga	accgcaacgt	tgaaggagcc	11880
actcagccgc	gggtttctgg	agtttaatga	gctaagcaca	tacgtcagaa	accattattg	11940
cgcggtcaaa	agtcgcctaa	ggcactatc	agctagcaaa	tatttcttgt	caaaaatgct	12000
ccactgacgt	tccataaatt	cccctcggt	tccaattaga	gtctcatatt	cactctcaat	12060
ccaaataatc	tgacccggat	ctcgagatcg	aattcccgcg	gccgcgaatt	cactagtggg	12120
tccccgggta	cggtcagtc	cttatgttac	gtcctgtaga	aaccccaacc	cgtgaaatca	12180
aaaaactcga	cggtcgtgtg	gcattcagtc	tggatcgcg	aaactgtgga	attgagcagc	12240
gttggtggga	aagcgcgtta	caagaaagcc	gggcaattgc	tgtgccaggc	agttttaacg	12300
atcagttcgc	cgatgcagat	attcgtaatt	atgtgggcaa	cgtctggtat	cagcgcgaag	12360
tctttatacc	gaaagggttg	gcaggccagc	gtatcgtgct	gcgtttcgat	gcggtcactc	12420
atttggtcaa	agtgtgggtc	aataatcagg	aagtgtgga	gcatacaggc	ggctatacgc	12480
catttgaagc	cgatgtcacg	ccgtatgtta	ttgccgggaa	aagtgtacgt	atcacagttt	12540
gtgtgaacaa	cgaactgaac	tggcagacta	tcccgcgggg	aatgggtgatt	accgacgaaa	12600
acgggaagaa	aaagcagtc	tacttccatg	aatttcttaa	ctacgcgggg	atccatcgca	12660
gcgtaatgct	ctacaccacg	ccgaacacct	gggtggacga	tatcacccgtg	gtgacgcgatg	12720
tcgcgcaaga	ctgtcaaccac	gcgtctgttg	actggcagg	gggtggccaat	gggtgatgtca	12780
cggttgaaact	gcgtgatg	gatcaaacag	tggttgcaac	tggacaaggc	accagcgagg	12840
ctttgcaagt	gggtgaatccg	cacctctggc	aacggggtga	agggttatctc	tatgaactgt	12900
acgtcacagc	caaaagccag	acagagtgtg	atatctaccc	gctgcgcgtc	ggcatccggt	12960
cagtgccagt	gaaggcgcaa	cagttcctga	tcaaccacaa	accgttctac	tttactggct	13020
ttggccgtca	tgaagatg	gatttgccgc	gcaaaggatt	cgataacgtg	ctgatgggtg	13080
acgatcacgc	attaatggac	tggattgggg	ccaactccta	ccgtacctcg	cattaccctt	13140
acgtgaaga	gatgtcgac	tgggcagatg	aacatggcat	cgtgggtgatt	gatgaaactg	13200
cagctgtcgg	ctttaacctc	tctttaggca	ttggtttcga	agcgggcaac	aagccgaaag	13260
aactgtacag	cgaagaggca	gtcaacgggg	aaactcagca	ggcgcaacta	caggcgatta	13320
aagagctgat	agcgcgtgac	aaaaaccacc	caagcgtggt	gatgtggagt	attgccaacg	13380
aaccggatac	ccgtccgcaa	gggtgcacgg	aatatttctc	gccactggcg	gaagcaacgc	13440
gtaaaactcga	tccgacgcgt	ccgatcacct	gcgtcaatgt	aatgttctgc	gacgtccaca	13500
ccgataccat	cagcgatctc	tttgatgtgc	tgtgcctgaa	ccgttattac	ggttggttatg	13560
tccaaagcgg	cgatttggaa	acggcagaga	aggtaactgga	aaaagaactt	ctggcctggc	13620
aggagaaaact	gcatacagccg	attatcatca	ccgaatacgg	cgtggatacg	ttagccgggg	13680
tgactcaat	gtacaccgac	atgtggagt	aagagtatca	gtgtgcatgg	ctggatatgt	13740
atcacgcgt	ctttgatcgc	gtcagcgccg	tcgtcggtga	acaggtatgg	aatttcgccc	13800
attttgcgac	ctcgcaaggc	atattgcgcg	ttggcggtaa	caagaagggg	atcttcaccc	13860
gcgaccgcaa	accgaagtcg	gcggcttttc	tgtgcacaaa	acgtggact	ggcatgaact	13920
tcggtgaaaa	accgcagcag	ggaggcaaac	aatgaatcaa	caactctcct	ggcgcaacct	13980
cgtcgggtac	agcctcgggg	attgcgtacc	gagctcgaat	ttccccgatc	gttcaaacat	14040
ttggcaataa	agtttcttaa	gattgaatcc	tgttgcgggt	cttgcatga	ttatcatata	14100
atttctgttg	aattcgttta	agcatgtaat	acattaacatg	taatgcatga	cgttatttat	14160
gagatgggtt	tttatgatta	gagtcgccga	attatacatt	taatacgcga	tagaaaacaa	14220
aatatagcgc	gcaaactagg	ataaattatc	gcgcgcgggtg	tcactatgt	tactagatcg	14280
ggaattcgat	atcaagcctg	atcactggcg	tcggtttaca	acgtcgtgac	tgggaaaacc	14340
ctggcggttac	ccaacttaat	cgccttgcag	cacatcccc	tttcgccagc	tggcgtaata	14400
gcgaagaggg	ccgcaccgat	cgcccttccc	aacagttgcg	cagcctgaat	ggcgaatgct	14460
agagcagcct	gagcttggat	cagattgtcg	tttccgcct	tcagtttaaa	ctatcagttg	14520

ttgacaggat atattggcgg gtaaacctaa gagaaaagag cgtttattag aataacggat 14580
atttaaaagg gcgtgaaaag gtttatccgt tcgtccattt gtatgtg 14627

<210> 110
<211> 9080
<212> DNA
<213> Artificial Sequence

<220>
<223> p18attBZeo(6XHS4)2eGFP Plasmid

<400> 110
cagttgcccg ccgggtcgcg cagggcggaac tccccccccc acggctgctc gccgatctcg 60
gtcatggccg gcccgaggcg gtcccgggaag ttctgtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggcccgcg caccacacacc caggccaggg tgttgctccg caccacctgg 180
tcctggaccg cgctgatgaa cagggtcacg tcgtcccggg ccacaccggc gaagtcgtcc 240
tccacgaagt cccgggagaaa cccgagccgg tcgggtccaga actcgaccgc tccggcgacg 300
tcgcgcgccg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgatatc gaattcctgc 420
agcccccgcg atccgctcac ggggacagcc ccccccaaaa gccccaggcg atgtaattac 480
gtccctcccc cgctaggggg cagcagcgag ccgccccggg ctccgctccg gtccggcgct 540
cccccccgat ccccgagccg gcagcgtgcg gggacagccc gggcacgggg aaggtggcac 600
gggatcgctt tcctctgaac gcttctcgct ctcttttgag cctgcagaca cctgggggat 660
acggggcgcg ggatccgctc acggggacag ccccccccca aagccccag ggatgtaatt 720
acgtccctcc ccgctagggg ggcagcagcg agccgcccgg ggctccgctc cggctccggcg 780
ctccccccgc atccccgagc cggcagcgtg cggggacagc ccgggcacgg ggaaggtggc 840
acgggatcgc tttcctctga acgcttctcg ctgctctttg agcctgcaga cactcggggg 900
atacggggcc gcggatccgc tcacggggac agcccccccc caaagcccc agggatgtaa 960
ttacgtccct cccccgctag ggggcagcag cgagccgccc ggggctccgc tccggtccgg 1020
cgctcccccc gcccccga gcgcggcagc agcggggaca gcccgggcac ggggaaggtg 1080
gcacgggata gctttcctct gaacgcttct cgctgctctt tgagcctgca gacacctggg 1140
ggatacgggg ccgcggtatc gctcacgggg acagcccccc cccaaagccc ccagggatgt 1200
aattacgtcc ctccccgct agggggcagc agcgagccgc cgggggctcc gctccggctc 1260
ggcgctcccc ccgcatcccc gagccggcag cgtgcccggg cagcccgggc acggggaagg 1320
tggcacggga tcgctttcct ctgaacgctt ctccgctgctc tttgagcctg cagacacctg 1380
ggggatacgg gcgcggggat cgctcacggg ggacagcccc cccccaaagc cccagggat 1440
gtaattacgt cctcccccg ctagggggca gcagcgagcc gcccggggct ccgctccggg 1500
ccggcgctcc ccccgcatcc ccgagccggc agcgtgcccgg gacagcccgg gcacggggaa 1560
gggtggcagcg gatcgcttcc ctctgaacgc tctctgctgc tctttgagcc tgcagacacc 1620
tgggggatac ggggcccggg atccgctcac ggggacagcc ccccccaaaa gccccaggcg 1680
atgtaattac gtccctcccc cgctaggggg cagcagcgag ccgccccggg ctccgctccg 1740
gtccggcgct ccccccgat ccccgagccg ccagcgtgcg gggacagccc gggcacgggg 1800
aaggtggcac gggatcgctt tcctctgaac gcttctcgct gctctttgag cctgcagaca 1860
cctgggggat acggggcggg ggatccacta gttattaata gtaatcaatt acgggtaaat ggcccgctg 1920
tagttcatag cccatatatg gagttccgcg ttacataact tacggtaaat ggcccgctg 1980
gctgaccgcc caacgacccc cgccatttga cgtcaataat gacgtatgtt cccatagtaa 2040
gcgcaatagg gactttccat tgacgtcaat ggggtggaata tttacggtaa actgcccact 2100
tggcagtaca tcaagtgtat catatgccaa gtacgcccc tattgacgtc aatgacggta 2160
aatggcccgcc ctggcattat gccagtagca tgaccttatg ggactttcct acttggcagt 2220
acatctacgt attagtcata gctattacca tgggtcgagg tgagccccac gttctgcttc 2280
actctcccca tctccccccc ctcccccccc ccaattttgt atttatttat tttttaatta 2340
ttttgtgcag cgatgggggg gggggggggg ggggcgcgcg ccaggcgggg cggggcgggg 2400
cgagggggcg ggcggggcga ggcggagagg tgccggcgga gccaatcaga ggcggcgct 2460
ccgaaaagtt ccttttatgg cgaggccggc gcggcgggcg ccctataaaa agcgaagcgc 2520
gcggcgggcg ggagtgcgtg cgttgccctc gccccgtgcc ccgctccgcg ccgcctcgcg 2580
ccgcccgcgc cggctctgac tgaccgcgtt agccccacag gtgagcgggc gggacggccc 2640
ttctcctccg ggtgttaatt agcgtttgtt ttaatgacgg ctcgtttctt ttctgtgct 2700
gcgtgaaagc cttaaagggg tccgggaggg ccctttgtgc ggggggggagc ggctcggggg 2760
gtgcgtgcgt gtgtgtgtgc gtggggagcg ccgcgtgccc cccgcgctgc cggcgcgctg 2820
tgagcgtgcg gggcgcgccg cggggctttg tcgctccgca gtgtgcgca ggggagcgcg 2880
gccggggggg gtgccccggt gtgccccggg gctgcgaggg gaacaaaggc tgcgtgcggg 2940
gtgtgtgcgt ggggggggta gcaggggggt tgggcgcggc ggtcgggctg taaccccccc 3000
ctgcaccccc ccccccgagt tgetgagcac ggcgcgctt cgggtgcggg gctccgtgcg 3060
gggcgtggcg cggggctcgc cgtgccgggc ggggggtggc ggcaggtggg ggtgccgggc 3120
ggggcggggc cgcctcgggc cggggagggc tccggggagg ggcgcggcg ccccgagcgc 3180
ccggcggtcg tcgagcgcg gcgagccgca gccattgctt tttatggtaa tcgtgcgaga 3240
gggcgcaggg acttctttg tcccaaatct ggccggagcc aaatctggga ggcgcgcgcg 3300

cacccccctct	agcggggcgcg	ggcgaagcgcg	tgcgggcgcgc	gcaggaagga	aatggggcggg	3360
gagggccttc	gtgcgtcgcc	gcgcgcgcgt	ccccctctcc	atctccagcc	tcgggggctgc	3420
cgcaggggga	cggctgcctt	cggggggggac	ggggcagggc	gggggttcggc	ttctggcgctg	3480
tgaccggcg	ctctagagcc	tctgtctaac	atgttcatgc	cttctctctt	ttcctacagc	3540
tcctgggcaa	cgtgctgggt	gttgtgctgt	ctcatcattt	tggcaaagaa	ttcgccacca	3600
tggtgagcaa	gggcgaggag	ctgttcaccg	gggtgggtgcc	catcctgggt	gagctgggacg	3660
gcgacgtaaa	cggccacaag	ttcagcgtgt	ccggcgaggg	cgagggcgat	gccacctacg	3720
gcaagctgac	cctgaagttc	atctgcacca	ccggcaagct	gcccgtgccc	tggcccaccc	3780
tcgtgaccac	cctgacctac	ggcgtgcagt	gcttcagccg	ctaccccgac	cacatgaagc	3840
agcagcactt	cttcaagtcc	gccatgcccc	aaggctacgt	ccaggagcgc	accatcttct	3900
tcaaggacga	cggcaactac	aagaccgcg	ccgaggtgaa	gttcgagggc	gacacctgtg	3960
tgaaccgcat	cgagctgaag	ggcatcgact	tcaaggagga	cggcaacatc	ctggggcaca	4020
agctggagta	caacttcaac	agccacaacg	tctatatcat	ggccgcacaag	cagaagaacg	4080
gcatcagggt	gaacttcaag	atccgccaca	acatcgagga	cggcagcggtg	cagctcgccg	4140
accactacca	gcagaacacc	cccatcggcg	acggccccgt	gctgctgccc	gacaaccact	4200
acctgagcac	ccagtccgcc	ctgagcaaa	cccccaacga	gaagcgcgat	cacatggtcc	4260
tgctggagtt	cgtgaccgcc	gcccggatca	ctctcgccat	ggacgagctg	tacaagtaag	4320
aattcactcc	tcaggtgcag	gctgcctatc	agaagggtgg	ggctgggtgtg	gccaatgccc	4380
tggtcacaaa	ataccactga	gatctttttc	cctctgccaa	aaattatggg	gacatcatga	4440
agccccttga	gcactctgact	tctggctaat	aaaggaaatt	tattttcatt	gcaatagtgt	4500
gttggaaatt	tttgtgtctc	tcactcgga	ggacatatgg	gagggcaaat	catttaaaac	4560
atcagaatga	gtatttgggt	tagagtttgg	caacatatgc	catatgctgg	ctgccatgaa	4620
caaagggtgg	tataaagagg	tcacagtag	atgaacacg	cccctgctgt	ccattcetta	4680
ttccatagaa	aagccttgac	ttgaggttag	atttttttta	tattttgttt	tgtgttattt	4740
ttttctttta	catccctaaa	attttccctta	catgttttac	tagccagatt	tttccctctc	4800
tcctgactac	tcccagtcac	agctgtccct	cttctcttat	gaagatccct	cgacctgcag	4860
cccagcttg	catgcctgca	ggtcgactct	agtgcatccc	ccgccccgta	tccccagggt	4920
gtctgcaggg	tcaaagagca	gcgagaagcg	ttcagaggaa	agcgatcccc	tgccaccttc	4980
cccgtgcccc	ggctgtcccc	gcacgctgcc	ggctcgggga	tgcgggggga	gcgcgggacc	5040
ggagcgggag	cccgggcggc	tcgctgctgc	cccctagcgg	gggaggggacg	taattacatc	5100
cctggggggt	ttgggggggg	gctgtccccg	tgagcggatc	cgcggccccc	tatccccccag	5160
gtgtctgacg	gctcaaagag	cagcgagaag	cgttcagagg	aaagcgatcc	cgtgccacct	5220
tcctcgctgc	cgggctgtcc	ccgcacgctc	ccggtcgggg	gatgcggggg	gagcgccgga	5280
ccggagcgga	gccccggggc	gctcgtgctc	gccccctagc	gggggagggga	cgtaattaca	5340
tccttggggg	ctttgggggg	gggctgtccc	cgtgagcgga	tccgcggccc	cgatatcccc	5400
aggtgtctgc	aggctcaaag	agcagcgaga	agcgttcaga	ggaaagcgat	cccgtgccac	5460
cttccccgtg	cccgggctgt	ccccgcacgc	tgccgggtcg	gggatgcggg	gggagcgccg	5520
gaccggagcg	gagccccggg	cggctcgtgc	ctgcccccta	gccccggggg	gacgtaatta	5580
catccctggg	ggctttgggg	gggggctgtc	cccgtagcgc	gatccgcggc	cccgtatccc	5640
ccaggtgtct	gcaggctcaa	agagcagcga	gaagcgttca	gaggaaagcg	atcccgtgcc	5700
accttccccg	tgccccgggt	gtccccgcac	gctgccggct	cggggatgcg	gggggagcgc	5760
cggaccggag	cgagcccccg	ggcggtctgc	tgctgcccc	tagcggggga	gggagcgta	5820
tacatccctg	ggggctttgg	gggggggctg	tccccgtgag	cggatccgcg	gccccgtatc	5880
ccccaggtgt	ctgcaggctc	aaagagcagc	gagaagcgtt	cagaggaaag	cgatcccgtg	5940
ccaccttccc	cgtgccccgg	ctgtccccgc	acgtgccccg	ctcgggggatg	cgggggggagc	6000
gcccgaaccg	agcggagccc	cgggcgggctc	gctgctgccc	cctagcgggg	gagggacgta	6060
attacatccc	tgggggcttt	gggggggggg	tgccccctgt	agcggatccg	cggccccgta	6120
tccccagggt	gtctgcaggg	tcaaagagca	gcgagaagcg	ttcagaggaa	agcgatcccc	6180
tgccaccttc	cccgtgcccc	ggctgtcccc	gcacgctgcc	ggctcgggga	tgcgggggga	6240
gcgcgggacc	ggagcgggag	cccggggcgg	tcgctgctgc	cccctagcgg	gggaggggacg	6300
taattacate	cctggggggt	ttgggggggg	gctgtccccg	tgagcgggatc	cgcgggggctg	6360
caggaattcg	taatcatggt	catagctgtt	tcctgtgtga	aattgttatc	cgctcacaat	6420
tccacacaac	atacagagccg	gaagcataaa	gtgtaaagcc	tgggggtgcct	aatgagtgag	6480
ctaactcaca	ttaattgcgt	tgcgctcact	gcccgttttc	cagtcgggaa	acctgtcggtg	6540
ccagctgcat	taatgaatcg	gccaacgcgc	ggggagaggc	ggtttgcgta	ttgggcgctc	6600
ttccgcttcc	tcgctcactg	actcgctgcg	ctcggtcggt	cggctgcggc	gagcggtatc	6660
agctcactca	aaggcggtaa	tacggttatc	cacagaatca	ggggataacg	caggaaagaa	6720
catgtgagca	aaaggccagc	aaaaggccag	gaaccgtaaa	aaggccgcgt	tgctggcggtt	6780
tttccatagg	ctccgcccc	ctgacgagca	tcacaaaaat	cagcgtcaa	gtcagaggtg	6840
gcgaaccocg	acaggactat	aaagatacca	ggcgtttccc	cctggaagct	ccctcggtgc	6900
ctctcctgtt	ccgacctgct	cgcttaccgg	atacctgtcc	gcctttctcc	cttcgggaag	6960
cgtggcgctt	tctcatagct	cacgctgtag	gtatctcagt	tcggtgtagg	tcggtcgctc	7020
caagctgggc	tgtgtgcacg	aaccccccg	tcagcccgac	cgctgcgcct	tatccggtaa	7080
ctatcgtctt	gagtcacaac	cggttaagaca	cgacttatcg	ccactggcag	cagccactgg	7140
taacaggatt	agcagagcga	ggtaggttag	cgggtgctaca	gagttcttga	agtggtggcc	7200
taactacgg	tacactagaa	ggacagtagt	tggtatctgc	gctctgctga	agccagttac	7260
cttcggaaaa	agagttggta	gctcttgatc	cggcaaacaa	accaccgctg	gtagcgggtg	7320

tttttttgggt	tgcaagcagc	agattacgcg	cagaaaaaaa	ggatctcaag	aagatccttt	7380
gatcttttct	acgggggtctg	acgctcagtg	gaacgaaaaac	tcacgttaag	ggatttttgg	7440
catgagatta	tcaaaaagga	tcttcaccta	gatcctttta	aattaaaaat	gaagttttta	7500
atcaatctaa	agtatatatg	agtaaacttg	gtctgacagt	taccaatgct	taatcagtga	7560
ggcacctatc	tcagcgatct	gtctatttgc	ttcatccata	gttgccctgac	tccccgtcgt	7620
gtagataaact	acgatacggg	agggctttacc	atctggcccc	agtgtcgcaa	tgataccgcg	7680
agacccacgc	tcaccgggctc	cagattttatc	agcaataaac	cagccagccg	gaagggccga	7740
gcgcagaagt	ggctcctgcaa	ctttatccgc	ctccatccag	tctattaatt	gttgccggga	7800
agctagagta	agtagttcgc	cagttaatag	tttgcgcaac	gttggttgcca	ttgctacagg	7860
catcgtgggtg	tcacgctcgt	cgtttgggtat	ggcttcattc	agctccgggt	cccaacgatc	7920
aaggcgagtt	acatgatccc	ccatgtttgtg	caaaaaagcg	gttagctcct	tcggctcctcc	7980
gatcgttgct	agaagtaagt	tggccgcagt	gttatcactc	atgggttatgg	cagcactgca	8040
taattctctt	actgtcatgc	catccgtaag	atgcttttct	gtgactgggtg	agtactcaac	8100
caagtcatct	tgagaatagt	gtatgcggcg	accgagttgc	tcttgcccgg	cgtcaatacg	8160
ggataataacc	gcagaacttt	gcagaacttt	aaaagtgtctc	atcattggaa	aacgtttcttc	8220
ggggcgaaaa	ctctcaagga	tcttaccgct	gttgagatcc	agttcgatgt	aaccctactcg	8280
tgcacccaac	tgatcttcag	catctttttac	tttcaccagc	gtttctgggt	gagcaaaaaac	8340
aggaagcgaa	aagcccgcaa	aaaaggggaat	aagggcgaca	cggaaatgtt	gaatactcat	8400
actcttctct	tttcaatatt	attgaagcat	ttatcaggggt	tattgtctca	tgagcgggata	8460
catatttgaa	tgtatttaga	aaaataaaca	aataggggtt	ccgcgcacat	ttccccgaaa	8520
agtgcacact	gacgtagtta	acaaaaaaaa	gcccgccgaa	gcgggcttta	ttaccaagcg	8580
aagcgcatt	gcgcattcag	gctgcgcaac	tgttgggaag	ggcgatcggt	gcgggcctct	8640
tcgctattac	gccagctggc	gaaaggggga	tgtgtcgcaa	ggcgattaag	ttgggtaacg	8700
ccagggtttt	cccagtcacg	acgttgtaaa	acgacggcca	gtccgtaata	cgactcactt	8760
aagcccttga	ctagagggtc	gacggtatac	agacatgata	agatacattg	atgagtttgg	8820
acaaaccaca	actagaatgc	agtgaaaaaa	atgctttatt	tgtgaaattt	gtgatgctat	8880
tgctttattt	gtaaccatta	taagctgcaa	taaacaagtt	ggggttggcg	aagaaactcca	8940
gcatgagatc	cccgcgctgg	aggatcatcc	agccggcgctc	ccggaaaacg	attccgaagc	9000
ccaacctttc	atagaaggcg	gcggtggaat	cgaatctctg	tagcacgtgt	cagtccctgct	9060
cctcggccac	gaagtgcacg					9080

<210> 111

<211> 4223

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attBBSRpolyA10 Plasmid

<400> 111

gttaactacg	tcagggtggca	cttttcgggg	aaatgtgcgc	ggaaccccta	tttgtttatt	60
tttctaaata	cattcaataa	tgtatccgct	catgagacaa	taaccctgat	aaatgcttca	120
ataaatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttattccctt	180
ttttgcggca	ttttgccttc	ctgtttttgc	tcaccagaaa	acgctgggtga	aagtaaaaga	240
tgctgaagat	cagttgggtg	cacgagtggtg	ttacatcgaa	ctggatctca	acagcggtaa	300
gatccttgag	agttttcgcc	ccgaagaacg	ttctccaatg	atgagcactt	ttaaagttct	360
gctatgtggc	gcggtattat	cccgtgttga	cgccgggcaa	gagcaactcg	gtcgcgcgat	420
acactattct	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	480
tggcatgaca	gtaagagaat	tatgcagtgc	tgccataacc	atgagtata	acactgcggc	540
caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	tgcaacaacat	600
gggggatcat	gtaactcgcc	ttgatcgttg	ggaacccggag	ctgaatgaag	ccataccaaa	660
cgacgagcgt	gacaccacga	tgccgttagc	aatggcaaca	acgttgcgca	aactattaac	720
tggcgaaacta	cttactctag	cttcccggca	acaattaata	gactggatgg	aggcggataa	780
agttgcagga	ccacttctgc	gctcggccct	tccggctggc	tggtttattg	ctgataaatc	840
tggagccgggt	gagcgtgggt	ctcgcgggtat	cattgcagca	ctggggccag	atggtaagcc	900
ctcccgtatc	gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	960
acagatcgct	gagatagggtg	cctcactgat	taagcattgg	taactgtcag	accaagttta	1020
ctcatatata	ctttagattg	ctttaccccg	gttgataatc	agaaaagccc	caaaaacagg	1080
aagattgtat	aagcaaatat	ttaaattgta	aacgttaata	ttttgttaaa	attcgcgtta	1140
aatttttgggt	aaatcagctc	attttttaac	caataggccg	aaatcggcaa	aatcccttat	1200
aaatcaaaag	aatagccgga	gatagggttg	agtgttggtc	cagtttgga	caagagtcga	1260
ctattaaaga	acgtggactc	caacgtcaaa	ggcgcaaaaa	ccgtctatca	ggcgcatggc	1320
ccactacgtg	aaccatcacc	caaatcaagt	tttttgggggt	cgaggtgccg	taaaagcata	1380
aatcggaacc	ctaaagggag	cccccgattt	agagcttgac	ggggaaaagcg	aacgtggcga	1440
gaaaggaagg	gaagaaagcg	aaaggagcgg	gcgctagggc	gctggcaagt	gtagcgggtca	1500
cgctgcgcgt	aaccaccaca	cccgcgcgcg	ttaatgcggc	gctacagggc	gcgtaaaagc	1560
atctagggtga	agatcctttt	tgataatctc	atgaccaaaa	tcccttaacg	tgagttttcg	1620

ttccactgag	cgtcagaccc	cgtagaaaag	atcaaaggat	cttcttgaga	tccttttttt	1680
ctgcgcgtaa	tctgctgctt	gcaaacacaaa	aaaccaccgc	taccagcggg	ggtttggttg	1740
ccggatcaag	agctaccaac	tctttttccg	aaggtaactg	gcttcagcag	agcgcagata	1800
ccaaatactg	ttcttctagt	gtagccgtag	ttaggccacc	acttcaagaa	ctctgtagca	1860
ccgcctacat	acctcgctct	gctaatacctg	ttaccagtg	ctgctgccag	tggcgataag	1920
tcgtgtctta	ccgggttgga	ctcaagacga	tagttaccgg	ataaggcgca	gcgggtcgggc	1980
tgaacggggg	gttcgtgcac	acagcccagc	ttggagcgaa	cgacctacac	cgaactgaga	2040
tacctacagc	gtgagctatg	agaaaagcgc	acgcttccc	aaggagagaaa	ggcggacagg	2100
tatccggtaa	gcggcagggg	cggaacagga	gagcgcacga	gggagcttcc	aggggggaaac	2160
gcctgggtatc	tttatagctc	tgctcggttt	cgccacctct	gacttgagcg	tcgatttttg	2220
tgatgtctcg	cagggggggc	gagcctatgg	aaaaacgcca	gcaacgcggc	ctttttacgg	2280
ttcctggcct	tttgctggcc	ttttgctcac	atgtaatgtg	agttagctca	ctcattaggc	2340
acccaggct	ttacacttta	tgcttccggc	tcgtatgttg	tgtggaattg	tgagcggata	2400
acaattttcac	acaggaaaca	gctatgacca	tgattacgcc	aagctacgta	atacgaactca	2460
ctagttggggc	ccgtgcaatt	gaagccggct	ggcgccaagc	ttctctgcag	gattgaagcc	2520
tgcttttttta	tactaacttg	agcgaaatct	ggatcaccat	gaaaacattt	aacatttctc	2580
ataaagatct	agaattagta	gaagtacgca	cagagaagat	tacaatgctt	tatgaggata	2640
ataaacatca	tgtagggagc	gcaattcgta	cgaaaacagg	agaaatcatt	tcggcagtag	2700
atattgaagc	gtatatagga	cgagtaactg	tttgtgcaga	agccattgcg	atttggtagt	2760
cagtttcgaa	tggacaaaag	gattttgaca	cgattgtagc	tgttagacac	ccttattctg	2820
acgaagtaga	ttagagttat	cgagtggtaa	gtccttggtg	tatgtgtagg	gagttgattt	2880
cagactatgc	accagattgt	tttgtgttaa	tagaaatgaa	tggcaagtta	gtcaaaacta	2940
cgattgaaga	actcattcca	ctcaaatata	cccgaaatta	aaagttttac	cataccaagc	3000
ttggctgctg	cctgaggctg	gacgacctcg	cgagtttcta	ccggcagtg	aaatccgtcg	3060
gcatccagga	aaccagcagc	ggctatccgc	gcatccatgc	ccccgaactg	caggagtggg	3120
gaggcacgat	ggccgctttg	gtccggatct	ttgtgaagga	accttacttc	tgtgggtgtga	3180
cataatttga	caaaactacct	acagagattt	aaagctctaa	ggtaaatata	aaatttttaa	3240
gtgtataatg	tgtaaaacta	ctgattctaa	ttgtttgtgt	attttagatt	ccaacctatg	3300
gaactgatga	atgggagcag	tggtggaatg	cctttaatga	ggaaaacctg	ttttgctcag	3360
aagaaatgcc	atctagtgat	gatgaggcta	ctgctgactc	tcaacattct	actcctccaa	3420
aaaagaagag	aaaggtagaa	gaccccaagg	actttccttc	agaattgcta	agttttttga	3480
gtcatgctgt	gtttagtaat	agaactcttg	ccttgcttgc	tattttacacc	acaaaggaaa	3540
ataaggaata	gctatacaag	aaaattatgg	aaaaatattc	tgtaaccttt	ataagtaggc	3600
ataacagtta	taatcataac	atactgtttt	ttcttactcc	acacaggcat	agagtgtctg	3660
ctattaataa	ctatgctcaa	aaattgtgtg	cctttagctt	tttaatttgt	aaaggggtta	3720
aatagcatca	tttgatgtat	agtgccttga	ctagagatca	taatcagcca	taccacattt	3780
gtagagggtt	tacttgcttt	aaaaaacctc	ccacacctcc	ccctgaacct	gaaacataaa	3840
atgaatgcaa	ttgttggtgt	taacttggtt	attgcagctt	ataatgggtt	caaataaagc	3900
aatagcatca	caaatttcac	aaataaagat	ccacgaattc	gctagcttcg	gccgtgacgc	3960
gtctccggat	gtacaggcat	gcgtcgaccc	tctagtcaag	gccttaagt	agtcgtatta	4020
cggactggcc	gtcgtttttac	aacgtcgtga	ctgggaaaac	cctggcggtt	cccaacttaa	4080
tcgccttgca	gcacatcccc	ctttcgccag	ctggcgtaat	agcgaagagg	cccgcaccga	4140
tcgccttcc	caacagttgc	gcagcctgaa	tggcgaaatg	cgcttcgctt	ggtaataaag	4200
ccgccttcgg	cgggcttttt	ttt				4223

<210> 112
 <211> 5855
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pCX-LamIntR Plasmid

<400> 112						
gtcgacattg	attattgact	agttattaat	agtaatcaat	tacgggggtca	ttagttcata	60
gcccataat	ggagttccgc	gttacataac	ttacggtaaa	tggcccgcc	ggctgaccgc	120
ccaacgaccc	ccgccattg	acgtcaataa	tgacgtatgt	tcccatagta	acgccaatag	180
ggactttcca	ttgacgtcaa	tgggtggact	atttacggta	aactgcccac	ttggcagtag	240
atcaagtgtg	tcatatgcc	agtacgcccc	ctattgacgt	caatgacggt	aaatggcccc	300
cctggcattg	tgcccagtag	atgaccttat	gggacatttc	tacttggcag	tacatctacg	360
tattatgcat	cggtattacc	atgggtcgag	gtgagcccca	cgttctgctt	cactctcccc	420
atctcccccc	cctccccacc	cccaattttg	tattttattta	tttttttaatt	attttgtgca	480
gcgatggggg	cggggggggg	gggggcgcgc	gccaggcggg	gcggggcggg	gcgagggggc	540
gggcggggcg	gtgcggcgag	gtgcggcgcg	agccaatcag	agcggcgcg	tccgaaagt	600
tcctttttatg	gcgagggcgc	ggcggcgggc	gccctataaa	aagcgaagcg	cgcgggcggg	660
gggagtcgct	gcgttgctt	cgccccgtgc	cccgcctccg	gccgcctcgc	gccgcgccgc	720
ccggctctga	ctgaccgcgt	tactcccaca	ggtgagcggg	cgggacggcc	cttctcctcc	780

gggctgtaat	tagcgcttgg	tttaaatgacg	gctcgtttct	tttctgtggc	tgcgtgaaag	840
ccttaaagg	ctccgggagg	gccctttgtg	cgggggggag	cggtcgggg	gggtcggtcg	900
tgtgtgtgtg	cgtggggagg	gcccgctgcg	gcccgcgctg	cccgcgcgct	gtgagcgctg	960
cgggcgcggc	gcggggcttt	gtgcgctccg	cgtgtgcgcg	aggggagcgc	ggcggggggc	1020
ggtgccccgc	ggtgcggggg	ggctgcgagg	ggaacaaagg	ctgctgctcg	gggtgtgtcg	1080
tgggggggtg	agcagggggg	gtgggcgcgg	cggtcgggct	gtaaccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cggcccggtt	tccgggtgcg	ggctccgtgc	ggggcggtgg	1200
gcggggctcg	cctgcccggg	cggggggtgg	tttataggtg	gggtgccggg	cggggcgggg	1260
ccgcctcggg	ccggggaggg	ctcgggggag	gggcgcggcg	gccccggagc	gccccgggct	1320
gtcgaggcgc	ggcgagccgc	agccattgcc	ttttatggta	atcgctgcgag	agggcgcgag	1380
gacttccttt	gtcccaaata	tggcggagcc	gaaatctggg	aggcgccgcc	gcacccccct	1440
tagcggggcg	gggcgaagcg	gtgcgggcgc	ggcaggaagg	aaatgggagg	ggaggggcct	1500
cgtgcgtcgc	cgcgcgcgcg	tccccctctc	catctccagc	ctcggggctg	ccgcaggggg	1560
acggctgcct	tccgggggga	cggggcaggg	cggggttcgg	cttctggcgt	gtgacccggg	1620
gctctagagc	ctctgctaac	catgttcctg	ccttcttctt	tttctacag	ctcctgggca	1680
acgtgctggg	tggtgtgctg	tctcatcatt	ttggcaaaga	attcatggga	agaaggcgaa	1740
gtcatgagcg	cgggatttta	ccccctaacc	tttatataag	aaacaatgga	tattactgct	1800
acagggaccc	aaggacgggt	aaagagtttg	gattagggag	agacaggcga	atcgcaatca	1860
ctgaagctat	acaggccaac	attgagttat	tttcaggaca	caaacacaag	cctctgacag	1920
cgagaatcaa	cagtgataat	tccgttacgt	tacattcatg	gcttgatcgc	tacgaaaaaa	1980
tccgtggccag	cagaggaatc	aagcagaaga	cactcataaa	ttacatgagc	aaaattaaag	2040
caataaggag	gggtctgcct	gatgctccac	ttgaagacat	caccacaaaa	gaaattgagg	2100
caatgctcaa	tggatgcata	gacgagggca	aggcggcgtc	agccaagtta	atcagatcaa	2160
cactgagcga	tgcattccga	gaggcaatag	ctgaaggcca	tataacaaca	aacctgtctg	2220
ctgccactcg	cgcagcaaaa	tctagagtaa	ggagatcaag	acttacggct	gacgaatacc	2280
tgaataatcta	tcaagcagca	gaatcatcac	catgttgggt	cagacttgca	atggaactgg	2340
ctgttggttac	cgggcaacga	gttgggtgatt	tatgcgaaat	gaagtgggtc	gatatcgtag	2400
atggatatct	ttatgtcgag	caaagcaaaa	caggcgtaaa	aattggccatc	ccaacagcat	2460
tgcataattga	tgcctctcga	atatcaatga	aggaacacact	tgataaatgc	aaagagattc	2520
ttggcggaga	aaccataatt	gcactactac	gtcgcgaaac	gctttcatcc	ggcacagtat	2580
caaggatatt	tatgcgcgca	cgaaaagcat	caggctctttc	cttcgaaggg	gatccgccta	2640
cctttcacga	gttgccaggt	ttgtctgcaa	gactctatga	gaagcagata	agcgataagt	2700
ttgtcaacaa	tcttctcggg	cataagtcgg	acaccatggc	atcacagtat	cgtgatgaca	2760
gaggcagggg	gtgggacaaa	attgaaatca	aataagaatt	cactcctcag	gtgcaggctg	2820
cctatcagaa	gggtgtggct	gggtgtggca	atgccctggc	tcacaaatac	cactgagatc	2880
tttttccctc	tgccaaaaat	tatggggaca	ctatgaagcc	ccttgagcat	ctgacttctg	2940
gctaataaag	gaaattttatt	ttcatatgca	tagtgtgttg	gaattttttg	tgtctctcac	3000
tccgaaggac	atatgggagg	gcaaatcatt	taaaacatca	gaatgagtat	ttgggtttaga	3060
gtttggcaac	atatgccata	tgctggctgc	catgaacaaa	gggtggctata	aagaggttat	3120
cagtatatga	aacagcccc	tgctgtccat	tccttattcc	atagaaaagc	cttgacttga	3180
gggttagattt	tttttatatt	ttgtttttgt	ttattttttt	ctttaacatc	cctaaaaatt	3240
tccttactag	ttttctcagc	cagatttttc	ctctctctct	gactactccc	agtcattagct	3300
gtccctcttc	tcttatgaag	atccctcgac	ctgcagccca	agcttggcgt	aatcatgggtc	3360
atagctgttt	cctgtgtgaa	attgttatcc	gctcacaatt	ccacacaaca	tacgagccgg	3420
aagcataaag	tgtaagccct	gggtgtgcct	atgagtgagc	taactcacat	taattgctgt	3480
gcgctcactg	cccgcttttc	agtcgggaaa	cctgtcgtgc	cagcggatcc	gcactctcaat	3540
tagtcagcaa	ccatagtccc	gccccctaact	ccgcccattc	cgcccctaac	tccgcccagt	3600
tccgcccatt	ctccgcccc	tggctgacta	atttttttta	tttatgcaga	ggccgaggcc	3660
gcctcggcct	ctgagctatt	ccagaagtag	tgaggaggct	tttttgagg	cctaggcctt	3720
tgcaaaaagc	taacttggtt	attgcagctt	ataatggtta	caaataaagc	aatagcatca	3780
caaatttcac	aaataaagca	tttttttcac	tgcatctctag	ttgtgggttg	tccaaactca	3840
tcaatgtatc	ttatcatgtc	tggatccgct	gcattaatga	atcggccaac	gcgcggggag	3900
aggcggtttg	cgtattgggc	gctcttccgc	ttcctcgctc	actgactcgc	tgcgctcggt	3960
cgttcggctg	cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	4020
atcaggggat	aacgcaggaa	agaacatgtg	agcaaaaagg	cagcaaaaagg	ccaggaaccg	4080
taaaaaggcc	gcgtttgctg	cgtttttcca	taggctccgc	ccccctgacg	agcatcacaa	4140
aaatcgacgc	tcaagtcaga	gggtggcgaaa	ccgcacagga	ctataaagat	accaggcgtt	4200
tccccctgga	agctccctcg	tgcgctctcc	tgctccgacc	ctgcccgtta	ccggatacct	4260
gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctcaa	tgctcacgct	gtaggtatct	4320
cagttcggtg	taggtcgttc	gctccaaagt	gggctgtgtg	cacgaacccc	ccgttcagcc	4380
cgaccgctgc	gccttatccg	gtaactatcg	tcttgagttc	aacccggtaa	gacacgactt	4440
atcgccactg	gcagcagcca	ctggtaacag	gattagcaga	gcgaggtatg	taggcggtgc	4500
tacagagttc	ttgaagtggg	ggcctaacta	cggctacact	agaaggacag	tatttggtat	4560
ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	4620
acaaaccacc	gctggtagcg	gtgggttttt	tgtttgcaag	cagcagatta	cgcgacagaa	4680
aaaaggatct	caagaagatc	ccttgatctt	tctacgggg	tctgacgctc	agtggaaacg	4740
aaactcacgt	taagggaattt	tggtcatgag	attatcaaaa	aggatcttca	cctagatcct	4800


```
tttaaattaa aaatgaagtt ttaaataaat ctaaagtata tatgagtaaa cttgggtctga 4860
cagttaccaa tgcttaataca gtgaggcacc tatctcagcg atctgtctat ttcgttcac 4920
catagttgcc tgaactccccg tcgtgtagat aactacgata cgggaggggct taccatctgg 4980
ccccagtgt gcaatgatac cgcgagaccc acgctcaccg gctccagatt tatcagcaat 5040
aaaccagcca gccggaagg ggcgagcgag aagtggctct gcaactttat ccgcctccat 5100
ccagtctatt aattggtgcc gggaagctag agtaagtagt tcgccagtta atagtttgcg 5160
caacgttggt gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc 5220
attcagctcc ggttcccaac gatcaaggcg agttacatga tccccatgt tgtgcaaaaa 5280
agcgggttagc tccttcggtc ctccgatcgt tgtcagaagt aagttggccg cagtgttatc 5340
actcatgggt atggcagcac tgcataatct tcttactgtc atgccatccg taagatgctt 5400
ttctgtgact ggtgagtact caaccaagtc attctgagaa tagtgtatgc ggcgaccgag 5460
ttgtctttgc cggcgctcaa tacgggataa taccgcgcca catagcagaa ctttaaaagt 5520
gctcatcatt ggaaaacgtt ctccggggcg aaaactctca aggatcttac cgctgttgag 5580
atccagttcg atgtaaccca ctcgtcacac caactgatct tcagcatctt ttactttcac 5640
cagcgtttct ggggtgagaa aaacaggaag gcaaaaaagg gaataatggc 5700
gacacggaaa tgttgaaata tcatactctt cctttttcaa tattattgaa gcatttatca 5760
gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata aacaaatagg 5820
ggttccgcgc acatttcccc gaaaagtgcc acctg 5855
```

<210> 113

<211> 4346

<212> DNA

<213> Artificial Sequence

<220>

<223> pSV40-193AttpsensePur Plasmid

<400> 113

```
ccggtgcccc caccatcccc tgacccacgc cctgacccc tcacaaggag acgaccttcc 60
atgaccgagt acaagcccac ggtgcgcctc gccacccgcg acgacgtccc ccgggcccga 120
cgcacccctcg ccgcgcgctt cgccgactac cccgccacgc gccacaccgt cgaccccggac 180
cgccacatcg agcgggtcac cgagctgcaa gaactcttcc tcacgcgcgt ccgggtccgac 240
atcggaagg tgtgggtcgc ggacgacggc gccgctggac cagcccgag 300
agcgtcgaag cggggggcgg gtctgcgcgag atcgcccgcg gcatggccga gttgagcgg 360
tcccggtgg ccgcgcgagca acagatggaa ggcctcctgg cgccgcaccc gcccaaggag 420
cccgctgggt tcttggccac cgtcggcgtc tcgcccgcac accaggggca ggggtctggg 480
agcgcgctcg tgctccccgg agtgaggcgc gccgagcgcg ccgggggtgcc cgccttctcg 540
gagacctccg tgccccgcaa cctccccctc tacgagcggc tcgggttcac cgtcacccgc 600
gacgtcgagg tgccgaagg accgcgcacc tgggtgatga cccgcaagcc cgggtcctga 660
cgcccccccc acgacccgca gcgcccgcac gaaaggagcg cacgacccca tgggtccgac 720
cgaagccgac ccggggcgcc ccgcgcgacc cagacccgcc cccgaggccc accgactcta 780
gaggatcata atcagcata ccacatttgt agaggtttta ctgtctttaa aaaacctccc 840
acacctcccc ctgaacctga aacataaaat gaatgcaatt gttgttgtaa acttgtttat 900
tgcagcttat aatgggtaca aataaagcaa tagcatcaca aatttcacaa ataaagcatt 960
tttttctctg cattctagtt gtggtttgtc caaactcatc aatgtatctt atcatgtctg 1020
gatccgcgcc ttaagcttag agtcgactgt ttaaacctgc aggcattgca 1080
gcttggcgta atcatgggtc tagctgtttc ctgtgtgaaa ttgttatccg ctcaaatc 1140
cacacaacat acgagccgga agcataaagt gtaaaagcct gggtgcctaa tgagtgcct 1200
aactcacatt aattgcgttg cgtcactgca ccgctttcca gtcgggaaac ctgctcggtc 1260
agctgcatta atgaatcggc caacgcgcgg ggagaggcgg tttgcgtatt gggcgctctt 1320
ccgcttcttc gctcactgac tcgctgcgct cggctcgttc gctgcggcga ggggtatcag 1380
ctcactcaaa ggccggtata cggttatcca cagaatcagg ggataacgca ggaaagaaca 1440
tgtgagcaaa aggcagcaaa aaggccagga accgtaaaaa ggccgcgttg ctggcgcttt 1500
tccatagggt ccgccccctt gacgagcacc ccaaaaaatc acgctcaagt cagagggtgc 1560
gaaacccgac aggactataa agataaccag cgtttcccc tggaagctcc ctcggtcgct 1620
ctcctgttcc gacctgccc cttaccggat acctgtccgc ctttctccct tcgggaagcg 1680
tggcgcttcc tcatagctca cgctgtaggt atctcagttc ggtgtaggtc gttcgctcca 1740
agctgggctg tgtgcacgaa cccccggttc agcccagacc ctgcgcctta tccggtaact 1800
atcgtcttga gtccaaaccc gtaagacacg acttatcgcc actggcagca gccactggta 1860
acaggattag cagagcgagg tatgtaggcg gtgctacaga gttcttgaag tgggtggccta 1920
actacggcta cactagaagg acagtatttg gtatctgcgc tctgctgaag ccagttaact 1980
tcggaaaaag agttggtagc tottgatccg gcaaacaaac caccgctggg agcgggtgggt 2040
tttttgtttg caagcagcag attacgcgca gaaaaaaagg atctcaagaa gatcctttga 2100
tctttcttac ggggtctgac gctcagtgga acgaaaaact acgttaaggg attttgggtc 2160
tgagattatc aaaaaggatc ttcacctaga tcctttttaa ttaaaaatga agtttttaaa 2220
caatctaaag tatatatgag taaacttggt ctgacagtta ccaatgctta atcagtgagg 2280
cacctatctc agcgatctgt ctatttcgtt catccatagt tgcctgactc cccgctcggt 2340
```

```

agataactac gatacgggag ggcttaccat ctggcccccag tgctgcaatg ataccgcgag 2400
accacgcctc accgggctcca gatttatcag caataaacca gccagccgga agggcccgagc 2460
gcagaagtgg tccctgcaact ttatccgctt ccatccagtc tattaattgt tgccgggaag 2520
ctagagtaag tagttcgcca gttaatagtt tgcgcacacgt tggtgccatt gctacaggca 2580
tcgtgggtgc acgctcgctcg tttgggtatgg cttcattcag ctccgggttc caacgatcaa 2640
ggcgagttac atgatcccc atgttggtgca aaaaagcggg tagtccttc ggtcctccga 2700
tcgttgctcag aagtaagttg gcgcagtggt tatcactcat ggttatggca gcaactgcata 2760
attctcttac tgtcatgcca tccgtaagat gcttttctgt gactggtgag tactcaacca 2820
agtcattctg agaatagtg atgcggcgac agagtgtctc ttgcccggcg tcaatacggg 2880
ataataccgc gccacatagc agaactttaa aagtgtctat cattggaaaa cgttcttcgg 2940
ggcgaaaaact ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg 3000
cacccaactg atcttcagca tcttttactt tcaccagcgt ttctgggtga gcaaaaaacag 3060
gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg gaaatgttga atactcatac 3120
tcttctcttt tcaatattat tgaagcattt atcaggggta ttgtctcatg agcggataca 3180
tatttgaatg tatttagaaa aataaacaata taggggttcc gcgcacattt ccccgaaaag 3240
tgccacctga cgtctaagaa accattatta tcatgacatt aacctataaa aataggcgta 3300
tcacgaggcc ctttcgtctc gcgcgtttcg gtgatgacgg tgaaaaacct tgacacatgc 3360
agctcccgga gacggtcaca gcttgtctgt aagcggatgc cgggagcaga caagcccgct 3420
agggcgcgct agcgggtgtt gcgggtgtc ggggctggct taactatgag gcatcagagc 3480
agattgtact gagagtgcac catatgcggt gtgaaatacc gcacagatgc gtaaggagaa 3540
aataccgcat caggcgccat tcgcctatta gcgcagctgg caaccagggg atgtgtctga aggcgattaa 3600
tgccggcctc ttgcctatta tcccagtcac gacgttgtaa aacgacggcc agtgatttcg 3660
gttgggtaac gtcgtgtcag ttaggtgtg gaaagtcacc aggtctccca gcaggcagaa 3720
agctgtggaa catgcatctc aatagtcag agcatgcatc tcaattagtc agcaaccata gtcccgcctc 3780
gtagcaaaag catgcatctc aatagtcag agcatgcatc ctaactccgc ccagtctccg ggcctctgag 3840
cagcaggcag aagtatgcaa agcatgcatc ctaactccgc ccagtctccg ggcctctgag 3900
taactccgcc catcccgcct ttttatttat ggaggctcgg gttgattcat atgtactgca 3960
gactaatttt aggtcttttt catctaaagta gttttttatg atactaagtt ggcattataa 4020
agtagtgagg aggtcttttt catctaaagta gttttttatg atactaagtt ggcattataa 4080
tcactaatac gttttttatg atactaagtt ggcattataa aataaaaatca 4140
tatgtagctc gttttttatg atactaagtt ggcattataa aataaaaatca 4200
gtttctcggt aacagggtcac tatcagtgcaa ggcgcg 4260
tggtgcaacg cctctctgggg ggccgcg 4320
cccactccct

```

<210> 114
 <211> 3166
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> p18attBZeo Plasmid

```

<400> 114
cagttgcccg cggggtcgcg cagggcggaac tccgcctccc acgggtgctc gccgatctcg 60
gtcatggccg gcccggaagg gtcccggaaag ttcgtggaca cgacctccga ccaactggcg 120
tacagctcgt ccaggcccgcg caccacacacc caggccaggg tggtgtccgg caccacctgg 180
tccctggaccg cgtgatgaa cccgagccgg tccgtcccga ccacaccggc gaagtgcgtc 240
tccacgaagt cccgggagaa cccgagccgg tccgtcccga actcgaccgc tccggcgagc 300
tcgcgcgcgg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgcatgc ctgcaggctc 420
actctagagg atccccgggt accgagctcg aattcgtaat catggtcata gctgtttcct 480
gtgtgaaatt gttatccgct cacaattcca cacaacatac gagccggaag cataaagtgt 540
aaagcctggg gtgcctaatt agtgagctaa ctcacattaa ttgctgtgag ctactgccc 600
gctttccagt cgggaaacct gtcgtgccag ctgcattaat gaatcggcca acgcgcgggg 660
agaggcggtt tgcgtattgg gcgctcttcc gcttccctgc tcaactgact cgtgcgctcg 720
gtcgttcggc tgccggcgagc ggtatcagct cactcaaaag cggtaatacg gttatccaca 780
gaatcagggg ataacgcagg aaagaacatg tgagcaaaa gcccagcaaa ggccaggaa 840
cgtaaaaagg ccgctgtgtg ggcgtttttc catagggctc gcccctctga cgagcatcac 900
aaaaatcgac gctcaagtca gaggtggcga aaccgcagag gactataaag ataccaggcg 960
tttccccctg gaagctccct cgtgcgctct cctgtccga cctgcccgt taccggatac 1020
ctgtccgctt ttctcccttc gggaagcgtg gcgctttctc atagctcacg ctgtaggtat 1080
ctcagttcgg tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc ccccggttcag 1140
cccgaccgct gcgccttata cgttaactat cgtcttgagt ccaaccgggt aagcacgcac 1200
ttatcgccac tggcagcagc cactggtaac aggtatagca gagcgaggta tgtaggcggt 1260
gtacagaggt tcttgaagtg gtggcctaac tacggctaca ctagaaggac agtatttgg 1320
atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc 1380

```

```
aaacaaacca cgcgtggttag cgggtggtttt tttggttgca agcagcagat tacgcgcaga 1440
aaaaaaggat ctcaagaaga tcttttgatc ttttctacgg ggtctgacgc tcagtggaac 1500
gaaaactcac gttaagggat tttgggtcatg agattatcaa aaaggatctt cacctagatc 1560
cttttaaat aaaaatgaag ttttaaatca atctaaagta tatatgagta aacttggtct 1620
gacagttacc aatgcttaat cagtggaggca cctatctcag cgatctgtct atttcgttca 1680
tccatagttg cctgactccc cgctcgttag ataactacga tacgggaggg cttaccatct 1740
ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca 1800
ataaaccagc cagccggaag ggccgagcgc agaagtgggtc ctgcaacttt atccgcctcc 1860
atccagtccta ttaattggtg ccgggaagct agagtaagta gttcgccagt taatagtttg 1920
cgcaacggtt ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggatgggt 1980
tcattcagct ccggttcccc acgatcaagg gtgtgcagaa gtaagttggc cgcagtgtta 2040
aaagcgggta gctccttcgg tccctcgatc tctcttactg tcatgccatc cgtaagatgc 2100
tactcatggg ttatggcagc actgcataat tcttctactg tcatgccatc cgtaagatgc 2160
ttttctgtga cgggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg 2220
agtgtctctt gcccgcgctc aataccgcgc cactatagcag aactttaaaa aactttaaaa 2280
gtgctcatca ttggaaaaacg ttcttcgggg cgaaaaactc caaggatctt accgctgttg 2340
agatccagtt cgatgttaac cactcgtgca ccaactgat cttcagcatc ttttactttc 2400
accagcggtt ctgggtgagc aaaaaacagg aggcataatg ccgcaaaaaa gggaataagg 2460
gcgacacgga aatggtgaat actcatactc ttcttttttc aatattattg aagcatttat 2520
cagggttatt gtctcatgag cggatacata tttgaatgta tttagaaaaa taaacaaata 2580
ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg tagttaacaa aaaaaagccc 2640
gccgaagcgg gctttattac caagcgaagc gccattcgcc attcaggctg cgcaactgtt 2700
gggaagggcg atcgggtcgg gcctcttcgc tattacgcca gctggcgaaa ggggggatgtg 2760
ctgcaaggcg attaatgttg gtaacgccag ggtttttcca gtacacgacg gtatacagac 2820
cggccagtcg gtaatacagc tcacttaagg ccttgactag agggctcgacg gaatgcagtg aaaaaaatgc 2880
atgataagat acattgatga gtttggacaa accacaacta gaatgcagtg aaaaaaatgc 2940
tttattttgt aaatttgtga tgctattgct ttattttaa ccatataag ctgcaataaa 3000
caagtggggg tgggcgaaga actccagcat gagatccccg cgctggagga tcatccagcc 3060
ggcgtcccg aaacgattc cgaagcccaa cctttcatag aaggcgggcg tggaatcgaa 3120
atctcgtagc acgtgtcagt cctgctctc ggccacgaag tgcacg 3160
```

<210> 115

<211> 7600

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeo3'6XHS4eGFP Plasmid

<400> 115

```
cagttgccgg cggggtcgcg cagggcggaac tcccgcctcc acggctgctc gccgatctcg 60
gtcatggcgg gcccggaagg gtcccggagc ttcgtggaca cgacctcga ccactcggg 120
tacagctcgt ccaggccgcg caccacacac cagggcaggg tggtgtccgg caccacctgg 180
tccctggaccg cgctgatgaa cagggtcacg tcgtcccggg ccacaccggc gaagtgcgtc 240
tccacgaagt cccgggagaa cccgagccgg tcggtccaga actcgaccgc tccggcgacg 300
tcgctcgcgg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgatcta gttattaata 420
gtaatcaatt acggggtcat tagttcatag cccatatatg gagttccgcg ttacataact 480
tacggtaaat ggcccgcctg gctgaccgcc caacgacccc cgccatttga cgtcaataat 540
gacgtatggt cccatagtaa ccccatagta actgcccact tggcagtaca gcccatttga ggggtggacta 600
tttacggtaa aatgacggta aatggcccgc ctggcattat gcccagtaca tgaccttatg 660
tattgacgtc acttggcagt acatctacgt tctctcccca cgtatggggc gggggggggg 720
tgagcccccac gttctgcttc ttttgtgcag cgagggggcg ggccggggcg ggcggagagg 780
atattattat tttttaatta ttttgtgcag cgagggggcg ggcggagagg cgagggcgcg ggcggcgcg 840
ccaggcgggg cggggcgggg ccgaaaagttt ccttttatgg ggttgccttc gcccggtgcc 900
gccaatcaga cggcgcgctt gggcgggggc cggctcgtcg ccgcccggcc cgtccacag 960
ccctataaaa agcgagcgcg cggcctcgcg cggcctcgcg cggcctcgcg cggcctcgcg 1020
gtgagcgggc gggcagggcc ttctcctcgg gctgaaagc gtgtgtgtgc gggcgcgcg 1080
ctcgtttctt tctgtgggt gtgctggggg tgagcgtcgt gcccggggcg gtgccccg 1140
ggggggggag ggctgggggg gtgctggggg tgagcgtcgt gcccggggcg gtgccccg 1200
cccgcgctgc cggcggggtt ggggagcgcg ggggagcgcg ggggagcgcg ggggagcgcg 1260
gtgtgcgcga ggggagcgcg ggggagcgcg ggggagcgcg ggggagcgcg ggggagcgcg 1320
gaacaaagcg tgcgtgcggg gtgcgtgcgg gtgcgtgcgg gtgcgtgcgg gtgcgtgcgg 1380
ggcggggctg taaccccccc cgtgcgtgcg gggcggtgcg gggcggtgcg gggcggtgcg 1440
cgggtgcggg gctcgtgcg 999cgtggcg cggggctcgc cgtgcggggc ggggggtggt 1500
ggcggggctg taaccccccc cgtgcgtgcg gggcggtgcg gggcggtgcg gggcggtgcg 1560
cgggtgcggg gctcgtgcg 999cgtggcg cggggctcgc cgtgcggggc ggggggtggt 1620
```

ggcaggtggg	ggtgcccggg	ggggcggggc	cgccctcggg	cggggagggc	tccggggagg	1680
ggcgccggcg	ccccgggagc	ccggcgggct	tccaggccgc	gcccagccgc	gccattgcct	1740
tttatggtaa	tccgtgcgaga	ggggcgaggc	acttcctttg	tcccaaatct	ggcggagccg	1800
aaatctggga	ggcgccgcgc	cacccctctt	agcgggcgcg	ggcgaagcgg	tgcggcgccg	1860
gcaggaagga	aatggggcgg	gagggccttc	gtgcgtcgcc	gcccgcgcgt	ccccttctcc	1920
atctccagcc	tccgggctgc	cgagggggga	cggtgccttc	cgggggggac	ggggcagggc	1980
gggggttcgg	ttctggcggt	tgaccggcgg	ctctagagcc	tctgctaacc	atgttcatgc	2040
cttcttcttt	ttcctacagc	tccctgggcaa	cgtgctgggt	gttgtgctgt	ctcatcattt	2100
tggcaaaagaa	ttcgccacca	tggtagagcaa	ggggcaggag	ctgttcaccg	gggtgggtcc	2160
catcctgggt	gagctgggac	gagacgtaaa	cgccacacaag	ttcagcgtgt	ccggcgaggg	2220
cgagggcgat	gccacctacg	gcaagctgac	cctgaagtct	atctgcacca	ccggcaagct	2280
gcccgtgccc	tggcccaccc	tccgtgaccac	cctgacctac	ggcgtgcagt	gcttcagccg	2340
ctaccccgac	cacatgaagc	agcacgactt	cttcaagtcc	gccatgcccg	aaggctacgt	2400
ccaggagcgc	accatcttct	tcaaggacga	cgccaactac	aagaccgcgc	ccgaggtgaa	2460
gttcgagggc	gacaccttgg	tgaaccggcg	cctgacgtga	ggcatcgact	tcaaggagga	2520
cggcaacatc	ctggggcaca	agctggagta	caactacaac	agccacaacg	tctatatcat	2580
ggccgcacaag	cagaagaacg	gcatcaagggt	gaacttcaag	atccgccaca	acatcgagga	2640
cggcagcggt	cagctcgccg	accactacca	ccagtccgcc	cccatcgggc	acggcccccgt	2700
gctgctgccc	gacaaccact	acctgagcac	cgtgaccgcc	ctgagcaaa	accccaacga	2760
gaagcgcgat	cacatgggtc	tgctggagtt	cctgagccgc	gcccgggatca	ctctcgggcat	2820
ggacgagctg	tacaagtaag	aattcactcc	tccaggtcag	gctgcctatc	agaaggtggg	2880
ggctgggtgt	gccaatgccc	tgggtcacia	ataccactga	gatctttttc	cctctgccaa	2940
aaattatggg	gacatcatga	agcccttga	gcatctgact	tctgggtaat	aaaggaaatt	3000
tattttcatt	gcaatagtgt	gttggaaatt	tttgtgtctc	tcaactcgga	ggacatatgg	3060
gagggcaaat	catttaaaac	atcagaatga	gtattttggt	tagagtttgg	caacatatgc	3120
catatgctgg	ctgccatgaa	caaagggtgg	tataaagagg	tcatcagtat	atgaaacagc	3180
cccctgctgt	ccattcctta	ttccatagaa	aagccttgac	ttgaggttag	atgtttttac	3240
tattttgttt	ttttctttta	ttttctttta	catccctaaa	attttccctt	catgtttttc	3300
tagccagatt	tttccctctc	tccctgactac	tccagtcact	agctgtccct	cttctcttat	3360
gaagatccct	cgacctgcag	cccaagcttg	catgcctgca	ggtcgactct	agtggatccc	3420
ccgcccggta	gtctgcaggc	gtctgcaggc	tcaaagagca	gcgagaagcg	ttcagaggaa	3480
agcgatcccc	tgccaccttc	cccgtgcccc	ggctgtcccc	gcacgctgcc	ggctcggggg	3540
tgcgggggga	gcgcgggacc	ggagcgggag	cccggggcgg	tccgtgctgc	cccctagcgg	3600
gggagggacg	taattacatc	cctggggggt	ttgggggggg	gctgtccccg	tgagcggatc	3660
cgcgggcccc	tatccccccg	gtgtctgcag	gctcaaagag	cagcgagaag	cgttcagagg	3720
aaagcggatc	cgtgccacct	tccccgtgcc	cgggctgtcc	ccgcacgctg	ccggctcggg	3780
gatgcggggg	gagcgccgga	ccggagcggg	gccccggggc	gctcgctgtg	gccccctagc	3840
gggggagggg	cgtaattaca	tccctggggg	ctttgggggg	gggctgtccc	cgtgagcggg	3900
tccgcggccc	cgtatcccc	aggtgtctgc	aggtcctaa	agcagcgaga	agcgttcaga	3960
ggaaagcgat	cccctggcct	cttccccgtg	cccggtgtgt	ccccgcacgc	tgccggctcg	4020
gggatgcggg	gggagcgccg	gaccggagcg	gagccccggg	cggctcgctg	ctgcccccta	4080
gcgggggagg	cagctaatta	catccctggg	ggctttgggg	gggggctgtc	cccgtgagcg	4140
gatccgcggc	cccgtatccc	ccagggtgtc	cgaggtcaca	agagcagcga	gaagcgttca	4200
gaggaaagcg	atcccggtgc	accttcccc	tgcccgggct	gtccccgcac	gctgcccggc	4260
cggggatgcg	gggggagcgc	cggaccggag	cggagccccg	ggcggtcgc	tgctgcccc	4320
tagcggggga	gggacgtaat	tacatccctg	gggcttttgg	gggggggctg	tccccgtgag	4380
cggatccgcg	gccccgtatc	ccccaggtgt	ctgcaggctc	aaagagcagc	gagaagcgtt	4440
cagaggaaag	cgatcccggt	ccaccttccc	cgtgccccgg	ctgtccccgc	acgctgcccg	4500
ctcggggatg	cggggggagc	gcccggaccg	agcggagccc	cgggcgggct	gctgctgccc	4560
cctagcgggg	gaggggacgt	attacatccc	tgggggcttt	gggggggggc	tgtccccctg	4620
agcggatccg	cggccccgta	tccccaggtt	gtctgcaggc	tcaaagagca	gcgagaagcg	4680
ttcagaggaa	agcgatcccc	tgccaccttc	cccgtgcccc	ggctgtcccc	gcacgctgcc	4740
ggctcggggg	tgccggggga	gcgcgggacc	ggagcgggac	cccgggcggc	tccgtgctgc	4800
cccctagcgg	gggagggagc	taattacatc	cctgggggct	ttgggggggg	gctgtccccg	4860
tgagcggatc	cgcggggctg	caggaattcg	taatcatggt	catagctgtt	tcctgtgtga	4920
aattgtttat	cgctcacaat	tccacacaac	atacagaccg	gaagcataaa	gtgtaaagcc	4980
tgggggtgct	aatgagtgag	ctaactcaca	ttaattgcgt	tgccgtcact	gcccgccttt	5040
cagtcgggaa	acctgtcgtg	ccagctgcac	taatgaatcg	gccaacgcgc	ggggagaggg	5100
ggtttgcgta	ttggggcgtc	ttccgcttcc	tccgtcactg	actcgctgcg	ctcggtcggt	5160
cggctgcggc	gagcgggtat	agctcactca	aaggcggtaa	tacgggttat	cacagaatca	5220
ggggataaac	caggaaaagaa	catgtgagca	aaaggccagc	aaaaggccag	gaaccgtaaa	5280
aaggcccgct	tgctggcggt	tttccatagg	ctccgcccc	ctgacgagca	tcacaaaaat	5340
cgacgctcaa	gtccagaggt	gcgaaaaccg	acaggactat	aaagatacca	ggcggtttcc	5400
cctggaagct	ccctcgtgcg	ctctcctgtt	cagacctgce	cgcttaccgg	atacctgtcc	5460
gccccttctc	cttcgggaag	cgtggcgctt	tctcatagct	cacgctgtag	gtatctcagt	5520
tccgtgtagg	tccgttcgct	caagctgggc	tgtgtgcacg	aacccccgt	tcagcccgac	5580
cgcgtgcgct	tatccggtaa	ctatcgtctt	gagtcacaac	cggtaagaca	cgacttatcg	5640

ccactggcag	cagccactgg	taacaggatt	agcagagcga	ggatatgtagg	cggtgctaca	5700
gagttcttga	agtgggtggcc	taactacggc	tacactagaa	ggacagtatt	tggtatctgc	5760
gctctgctga	agccagttac	cttcggaaaa	agagttggta	gctcttgatc	cgggcaaaaa	5820
accaccgctg	gtagcgggtg	tttttttgtt	tgcaagcagc	agattacgcg	cagaaaaaaa	5880
ggatctcaag	aagatccttt	gatcttttct	acggggtctg	acgctcagtg	gaacgaaaaa	5940
tcacgttaag	ggatttttgg	catgagatta	tcaaaaagga	tcttcaccta	gatcctttta	6000
aattaaaaat	gaagttttta	atcaatctaa	agtatatatg	agtaaacttg	gtctgacagt	6060
taccaatgct	taatcagtg	ggcacctatc	tcagcgatct	gtctatttctg	ttcatccata	6120
ggtgcctgac	tccccgtcgt	gtagataaact	acgatacggg	agggcttacc	atctggcccc	6180
agtgtgcaa	tgataccgcg	agacccacgc	tcaccggctc	cagattttatc	agcaataaac	6240
cagccagccg	gaagggccga	gcgacagaag	gggtcctgcaa	ctttatccgc	ctccatccag	6300
tctattaatt	ggtgcgggga	agctagagta	agtagttcgc	cagttaatag	tttgcgcaac	6360
ggtgttgcca	ttgctacagg	catcgtgggtg	tcacgctcgt	cgtttgggtat	ggcttcattc	6420
agctccgggt	cccaacgata	aaggcgagtt	acatgatccc	ccatggttg	caaaaaagcg	6480
ggttagctcct	toggctctcc	gatcgttgtc	agaagtaagt	tggccgcagt	gttatcactc	6540
atggttatgg	cagcactgca	taattctctt	actgtcatgc	catccgtaag	atgcttttct	6600
gtgactgggtg	agtactcaac	caagtcattc	tgagaatagt	gtatgccccg	accgagttgc	6660
tcttgccccg	cgtcaataacg	ggataatacc	gcgccacata	gcagaacttt	aaaagtgtctc	6720
atcattggaa	aacgttcttc	ggggcgaaaa	ctctcaagga	tcttaccgct	gttgagatcc	6780
agttcgatgt	aacccactcg	tgcacccaac	tgatcttcag	catcttttac	tttcaccagc	6840
gtttctgggt	gagcaaaaaa	aggaaggcaa	aatgcgcgca	aaaagggaat	aaggcgcaac	6900
cggaaatggt	gaatactcat	actcttctct	tttcaatatt	attgaagcat	ttatcagggt	6960
tattgtctca	tgagcgggata	catatttgaa	tgtattttaga	aaaataaaca	aataggggtt	7020
ccgcgcacat	ttccccgaaa	agtgccacct	gacgtagtta	acaaaaaaaa	gcccgcgcaa	7080
gcgggcttta	ttaccaagcg	aagcgccatt	cgccattcag	gctgcgcaac	tggtgggaag	7140
ggcgatcggt	gcgggcctct	tgcctattac	gccagctggc	gaaaggggga	tggtgctgcaa	7200
ggcgattaag	ttgggttaacg	ccagggtttt	cccagtcacg	acgttgtaaa	acgacggcca	7260
gtccgtaata	cgactcactt	aaggccttga	ctagaggggtc	gacggtatac	agacatgata	7320
agatacattg	atgagtttgg	acaaaccaca	actagaatgc	agtgaaaaaa	atgctttatt	7380
tgtgaaatgt	gtgatgctat	tgttttattt	gtaaccatta	taagctgcaa	taacaagatt	7440
ggggtgggcg	aagaactcca	gcatgagatc	cccgcgctgg	aggatcatcc	agccggcgctc	7500
ccggaaaaacg	attccgaagc	ccaacctttc	atagaaggcg	gcggtggaat	cgaatctctg	7560
tagcacgtgt	cagtcctgct	cctcggccac	gaagtgcacg			7600

<210> 116
<211> 7631
<212> DNA
<213> Artificial Sequence

<220>
<223> p18attBZeo5'6XHS4eGFP Plasmid

<400> 116						
cagttgcccg	ccgggtcgcg	cagggcgaac	tcccgcctcc	acggctgctc	gccgatctcg	60
gtcatggccg	gcccggaggc	gtcccgaag	ttcgtggaca	cgacctccga	ccactcggcg	120
tacagctcgt	ccaggccgcg	caccacaccc	caggccaggg	tggtgtccgg	caccacctgg	180
tcctggaccg	cgctgatgaa	cagggtcacg	tcgtcccggg	ccacaccggc	gaagtcgtcc	240
tccacgaagt	cccgggagaa	cccagccggg	tcgggtccaga	actcgaccgc	tccggcgacg	300
tcgcgcgcgg	tgagcaccgg	aacggcactg	gtcaacttgg	ccatggatcc	agatttcgct	360
caagttagta	taaaaaagca	ggcttcaatc	ctgcagagaa	gcttgatata	gaattcctgc	420
agcccccgcg	atccgctcac	ggggacagcc	cccccccaaa	gccccagggg	atgtaattac	480
gtccctcccc	cgctaggggg	cagcagcgag	ccgcgccggg	ctccgctccg	gtccggcgct	540
ccccccgcat	ccccgagccg	gcagcgtgcg	gggacagccc	gggcacgggg	aaggtggcac	600
gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	cctgggggat	660
acggggcccg	ggatccgctc	acggggacag	ccccccccca	aagccccag	ggatgtaatt	720
acgtccctcc	cccgttaggg	ggcagcagcg	agccgcgccg	ggctccgctc	cggtccggcg	780
ctccccccgc	atccccgagc	cggcagcgtg	cggggacagc	ccgggacagg	ggaaggtggc	840
acgggatcgc	tttctcttga	acgttctcgt	ctgctttttg	agcctgcaga	cacctggggg	900
atacggggcc	gcggatccgc	tcacggggac	agcccccccc	caaagccccc	agggatgtaa	960
ttacgtccct	cccccgctag	ggggcagcag	cgagccgccc	ggggctccgc	tccgggtccg	1020
cgctcccccc	gcataccccga	gccggcagcg	tcgggggaca	gcccgggcac	ggggaagggtg	1080
gcacgggata	gctttctctc	gaacgcttct	cgctgctctt	tgagcctgca	gacacctggg	1140
ggatacgggg	ccgcggatcc	gctcacgggg	acagcccccc	cccaaagccc	ccagggatgt	1200
aattacgtcc	ctcccccgct	agggggcagc	acgcagccgc	ccggggctcc	gctccggctc	1260
ggcgctcccc	ccgcctcccc	gagccggcag	cgtgcgggga	cagcccgggc	acggggaagg	1320
tggcacggga	tcgcttttct	ctgaacgctt	ctcgtgcttc	tttgagcctg	cagacacctg	1380
ggggatacgg	ggccgcggat	ccgctcacgg	ggacagcccc	cccccaaagc	ccccagggat	1440

gtaattacgt	ccctcccccg	ctaggggggca	gcagcgagcc	gccccggggct	ccgctccgggt	1500
ccggcgctcc	ccccgcattc	ccgagccggc	agcgtgcggg	gacagccccg	gcacgggggaa	1560
gggtggcacgg	gacgccttcc	ctctgaacgc	ttctcgctgc	tcttttgagcc	tgacagacacc	1620
tggggggatac	ggggcccgcg	atccgctcac	ggggacagcc	cccccccaaaa	gccccccagg	1680
atgtaattac	gtccctcccc	cgctaggggg	cagcagcgag	ccgccccgggg	ctccgctccg	1740
gtccggcgct	ccccccgcac	ccccgagccg	gcagcgtgcg	gggacagccc	gggcacgggg	1800
aagggtggcac	gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	1860
cctgggggat	acggggcggg	ggatccacta	gttattaata	gtaatcaatt	acgggggtcat	1920
tagttcatag	cccatatatg	gagttccgcg	ttacataact	tacggtaaat	ggccccgctg	1980
gctgaccgcc	caacgacccc	cgcccatgga	cgtaataaat	gacgtatggt	cccatagtaa	2040
cgccaatagg	gactttccat	tgacgtcaat	gggtggacta	tttacggtaa	actgcccact	2100
tggcagtaca	tcaagtgtat	catatgcaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggccccg	ctggcattat	gcccagtaca	tgacctatg	ggactttcct	acttggcagt	2220
acatctacgt	attagtcac	gctattacca	tgggtcgagg	tgagccccac	gttctgcttc	2280
actctcccc	tctccccccc	ctcccccccc	ccaattttgt	atttatttat	tttttaatta	2340
ttttgtgcag	cgatgggggg	gggggggggg	ggggcgcgcg	ccaggcgggg	cggggcgggg	2400
cgagggcgcg	ggcgggggcg	ggcgagagg	tgccggcgga	gccaatcaga	gcggcgcgct	2460
ccgaaagttt	ccttttatgg	cgaggcggg	gcggcgggcg	ccctataaaa	agcgaagcgc	2520
gcggcgggcg	ggagtgcctg	cgttgccttc	gccccgtgcc	ccgctccgcg	ccgcctcgcg	2580
ccgccccgcc	cggctctgac	tgaccgcggt	actcccacag	gtgagcgggc	gggacggccc	2640
ttctcctccg	ggctgtaatt	agcgcctggg	ttaatgacgg	ctcgtttctt	ttctgtggct	2700
cgtgaaaagc	cttaaaaggg	tccgggaggg	ccctttgtgc	ggggggggagc	ggctcggggg	2760
gtgctgctgc	gtgtgtgtgc	gtggggagcg	ccgcgtgctg	cccgcgctgc	ccggcggtcg	2820
tgagcgctgc	gggcgcgggc	cggggctttg	tgccgtccgc	gtgtgcgcga	ggggagcgcg	2880
gcccggggcg	gtgccccgcg	gtgcgggggg	gctgcgaggg	gaacaaaggg	tgctgctggg	2940
gtgtgtgctg	gggggggtga	gcaggggggt	tgggcgcggc	ggtcgggctg	taaccccccc	3000
ctgcaccccc	ctggcccgagt	tgctgagcac	ggccccggct	cggtgctggg	gctccgtgcg	3060
gggctgctgc	cgggctcgcg	cgtgcggggc	cggtgctggc	ggcaggtggg	ggtgccccgg	3120
ggggcggggg	cgcctcgggc	cggggagggc	tcgggggagg	ggcgcgcgcg	ccccggagcg	3180
ccggcggtcg	tcgagggcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcgaga	3240
ggggcgagg	acttcccttg	tccccaaact	ggcgagccgc	aaatctggga	ggcgccgcg	3300
cacccccctc	agcggggcg	ggcgaagcgg	tgccggcgcc	gcaggaagga	aatggggcg	3360
gagggccttc	gtgctgcgcc	gcgcgcgcgt	cccccttctc	atctccagcc	tcggggctgc	3420
cgcagggggg	cggggggggc	cggggggggg	gggggggggg	gggggttcgg	ttctggcggt	3480
tgaccggcg	ctctagagcc	tctgctaacc	atgttcatgc	cttcttcttt	ttcctacaga	3540
tcctgggcaa	cgtgctggtt	gttgtgctgt	ctcatcattt	tggcaaagaa	ttcgccacca	3600
tggttagcaa	ggcgagggag	ctgttcacgc	gggtgggtgc	catcctgggt	gagctggagc	3660
gcgacgtaaa	cggccacaag	ttcagcgtgt	ccggcgaggg	cgagggcgat	gccacctacg	3720
gcaagctgac	cctgaagttc	atctgcacca	ccggcaagct	gcccgtgccc	tggcccacc	3780
tcgtgaccac	cttgacctac	ggcgtgcagt	gcttcagccg	ctaccccgac	cacatgaagc	3840
agcacgactt	cttcaagtc	gccatgccc	aaggctacgt	ccaggagcgc	accatcttct	3900
tcaaggacga	cggcaactac	aagaccgcg	ccgaggtgaa	gttcgagggc	gacaccttgg	3960
tgaaccgcgt	cgagctgaag	ggcatcgact	tcaaggagga	cggcaacatc	ctggggcaca	4020
agctggagta	caactacaac	agccacaac	tctatatcat	ggccgacaag	cagaagaacg	4080
gcatcaaggt	gaacttcaag	atccgccaca	acatcgagga	cggcagcggt	cagctcgccg	4140
accactacca	ccagaacacc	cccatcgccg	acggccccgt	gctgctgccc	gacaaccact	4200
acctgagcac	ccagtccgcc	ctgagcaaag	accccaacga	gaagcgcgat	cacatgggtc	4260
tgctggagtt	cgtgaccgcc	gcccgggatc	ctctcgccat	ggacgagctg	tacaagtaag	4320
aattcactcc	tcaggtgcag	gctgcctatc	agaaggtggt	ggctgggtgtg	gccaatgccc	4380
tggtctacaa	ataccactga	gatctttttc	cctctgccaa	aaattatggg	gacatcatga	4440
agccccctga	gcactctgact	tctggctaatt	aaaggaaatt	tatttttcatt	gcaatagtgt	4500
gttggaaatt	tttgtgtctc	tcactcgga	ggacatatgg	gagggcaaat	cattttaaacc	4560
atcagaatga	gtatttgggt	tagagtgttg	caacatatgc	catatgctgg	ctgccatgaa	4620
caaaggtggc	tataaagagg	tcacagtgat	atgaaacagc	cccctgctgt	ccattcctta	4680
ttccatagaa	aagccctgac	ttgaggttag	atttttttta	tattttgttt	tgtgttattt	4740
ttttcttttaa	catccctaaa	attttcctta	catgtttttc	tagccagatt	tttctcctc	4800
tcctgactac	tcccagtcac	agctgtccct	cttctcttat	gaagatccct	cgacctgcag	4860
cccaagcttg	ctgcactgca	ggtcgactct	agaggatccc	cgggtaccga	gctcgaattc	4920
gtaatcatgg	tcatagtgtg	ttcctgtgtg	aaattgttat	ccgctcacia	ttccacacaa	4980
catacgagcc	ggaagcataa	agtgtaaagc	tggggggtgc	taatgagtga	gctaactcac	5040
attaattgog	ttgcgctcac	tgcccgtttt	ccagtcggga	aacctgtcgt	gccagtgca	5100
ttaatgaatc	ggccaacgcg	cggggagagg	cggtttgctg	attgggcgct	cttccgcttc	5160
ctcgctcact	gactcgctgc	gctcggtcgt	tcgggtgcgg	cgagcggtat	cagctcactc	5220
aaaggcggtg	atacggttat	ccacagaatc	aggggataac	gcaggaaaga	acatgtgagc	5280
aaaaggccag	caaaaaggcca	ggaaccgtaa	aaaggccgcg	ttgctggcgt	ttttccatag	5340
gctccgcccc	cctgacgagc	atcacaaaaa	tcgacgctca	agtcagaggt	ggcgaaaccc	5400
gacaggacta	taaagatacc	aggcgtttcc	ccctggaagc	tccctcgctg	gctctcctgt	5460

```

tccgaccctg cccgcttacgc gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct 5520
ttctcatagc tcacgctgta ggtatctcag ttccggtgtag gtcgttcgct ccaagctggg 5580
ctgtgtgcac gaaccccccg ttcagcccgga ccgctgcgcc ttatccggta actatcgctc 5640
tgagtccaac ccggtaagac acgacttatc gccactggca gcagccactg gtaacaggat 5700
tagcagagcg aggtatgtag gcggtgctac agagttcttg aagtgggtggc ctaactacgg 5760
ctacactaga aggacagtat ttggtatctg cgctctgctg aagccagtta ccttcggaaa 5820
aagagttggt agctcttgat ccggcaaaaa aaccaccgct ggtagcggtg gtttttttgt 5880
ttgcaagcag cagattacgc gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc 5940
tacgggggtc atcttcacct agatcctttt aaattaaaaa tgaagtttta aatcaatcta 6000
atcaaaaagg gagtaaacct ggtctgacag ttaccaatgc ttaatcagtg aggcacctat 6120
aagtatatat ggttaaacct cctccatcca catctggccc cagcactgca atgataccgc gagaccacg 6240
ctcagcgatc gagggccttac ccagatttat cagcaataaa ccagccagcc ggaaggggccc 6300
tacgatacgg actttatccg cctccatcca gtttgcgcaa cgttggtgcc attgctacag gcacgtggg 6420
ctcaccgggt ccagatttat cgttttggtt tggcttcatt cagctccggt tcccaacgat caaggcgagt 6480
aagtagttcg ccagtttaata tggcctcgag ggttagctcc ttccggtcctc cgatcgttgt 6540
gtcacgctcg ccatccgtaa gatgcttttc gacccagttg gcttgcgcc gcagcactgc ataattctct 6600
tacatgatcc ttggcgcgag gattatcact catgggttatg tgtgactggg gagtactcaa ccaagtcatt 6660
cagaagtaag ccacccgtaa gacgaaactt taaaagtgtc catcattgga aaacgttctt cggggcgaaa 6720
tactgtcatg actctacccg atcttaccgc tggttgagatc cagttcgatg taacccactc gtgcacccaa 6780
ctgagaatag tgtatgcggc gaccgagttg agcgaactt tggcggcggc cgtttctggg tgagcaaaaa caggaaggca 6900
cgccgacatc agcagaactt agcgggcttt acggaatgt tgaatactca tactcttctc 6960
actctcaagg atcttaccgc tggttgagatc cgtttctggg atgagcggat acatatttga 7020
ctgatcttca gcatctttta aaaaagggaa tattgaagca tttatcaggg tccgcgcaca ttccccgaa aagtgccac 7080
aaatgcccga aaaaataaac aacaaaaaaa ggctgcgcaa ataccgaagc gaagcgccat 7140
ttttcaatat atgtatttag agcgttgtaa agcccgccga ctgttgggaa gggcgatcgg tgcgggcctc ttcgctatta 7200
tgacgtcac cgccagctgg cgaaaagggg gacgttgtaa aacgacggc atgcccgaat attaccaagc ggcagggttt 7260
tcgccattca ggctgcgcaa atgtgctgca agcgagttg agcgactact gttgggtaac gccagggttt 7320
cgccagctgg gacgttgtaa cagacatgat agtccgtaat gatgagtttg taaggccttg 7380
actagagggt cagtgaaaaa aatgctttat ttgtgaaat tgtgatgcta ttgctttat 7440
aactagaatg ataagctgca ataacaagt tggggtgggc agcactgagat agcatgagat 7500
tgtaaccatt ccccgcgctg gaggatcatc cagccggcgt cccggaaaac gattccgaag cccaaccttt 7560
cccccgcgct catagaaggc ggcggtggaa tcgaaatctc gtagcacgtg tcagtcctgc tcctcggcca 7620
cgaagtgcac g 7631

```

<210> 117

<211> 4615

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeo6XHS4 Plasmid

<400> 117

```

cagttgcccg cccgggtcgcg cagggcgaa ccccccccc acggctgctc gccgatctcg 60
gtcatggccg gcccggaggc gtcccggaa ttcgtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggccgcg caccacacac caggccaggg tggtgtccgg caccacctgg 180
tcctggaccg cgctgatgaa cagggtcacg tcgtcccggg ccacaccggc gaagtgcgtc 240
tccacgaagt cccgggagaa cccgagccgg tccgtccaga actcgaccgc tccggcgacg 300
tcgcgcgcgg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagtttagt taaaaaagca ggcttcaatc ctgcagagaa gcttgcatgc ctgcaggctg 420
actctagtgg atcccccgcc ccgtatcccc caggtgtctg caggctcaaa gagcagcgag 480
aagcgttcag aggaaagcga tcccgtgcca ccttccccgt gcccggtctg tccccgcacg 540
ctgccggctc ggggatgcgg ggggagcgcc ggaccggagc ggagccccgg ggcgctcgct 600
gctgccccct agcgggggag ggaacttaatt acatccctgg gggctttggg ggggggctgt 660
ccccgtgagc ggatccgcgg ccccgatatc cccaggtgtc cccaggtgtc tgcaggctca aagagcagcg 720
agaagcgttc agaggaaagc gatcccgctg cacttcccc gtcggcgggc gtgccccggg tgtccccgca 780
cgctgcccgc tcggggatgc ggggggagcg cccgacggga gcgagcccc gggcgggctg 840
ctgctgcccc ctagcggggg agggacgtaa ttacatccct ggggggcttg ggggggggct 900
gtccccgtga cgggatccgc ggccccgtat cccccagggt tctgacaggc caaagagcag 960
cgagaagcgt tcagaggaaa gcgatcccg gccacctcc ccgtgcccc gctgtcccc 1020
cacgtgcgcg gctcggggat gcggggggag gcccgagacc gaggcggagc ccgggcggct 1080
cgctgctgcc ccctagcggg ggaggggacgt aattacatcc ctgggggctt tggggggggg 1140
ctgtccccgt gagcggatcc gcggccccgt atcccccagg tgtctgcagg ctcaaagagc 1200

```

```

agcgagaagc gttcagagga aagcgatccc gtgccacctt ccccggtgcc gggctgtccc 1260
cgcacgctgc cgggtcgggg atgccccggg agcgccggag cggagcggag ccccgggcgg 1320
ctcgctgctg cccctagcgg ggggagggac gtaattacat ccccgggggc tttggggggg 1380
ggctgtcccc gtgagcggat ccgcggcccc gtatccccca ggtgtctgca gggtcaaaga 1440
gcagcgagaa gcgttcagag gaaagcgatc ccgtgccacc ttccccgtgc ccgggctgtc 1500
cccgacgctg gccggctcgg ggatgcgggg ggagcgccgg accggagcgg agccccgggc 1560
ggctcgctgc tgccccctag cgggggaggg acgtaattac atccctgggg gctttggggg 1620
ggggctgtcc ccgtgagcgg atccgcggcc ccgtatcccc cagggtgtctg cagggtcaaa 1680
gagcagcgag aagcgttcag aggaaagcga tcccgtgcc cctccccgt gccgggctg 1740
tccccgcacg ctgccggctc ggggatgcgg ggggagcgcc ggaccggagc ggagccccgg 1800
gccccctcgt gctgccccct agcggggggg ggacgtaatt acatccctgg gggctttggg 1860
ggggggctgt ccccgtagc ggatcccgcg ggctgcagga attcgtaatc atgggtcatag 1920
ctgtttcctg tgtgaaattg ttatccgctc acaattccac acaacatacg agccggaagc 1980
ataaagtgtg aagcctgggg tgccctaatg gtgagctaac tcacattaat tgcgtttgcgc 2040
tactgcccg ctttccagtc gggaaacctg gctgtccagc tgcattaatg aatcgggcaa 2100
cgcgcgggga gaggcgggtt gcgtattggg gcgtcttccg cttcctcgct cactgactcg 2160
ctgcgctcgg tcgttcgggt gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg 2220
ttatccacag aatcagggga taacgcagga gcgtaactgt gagcaaaagg ccagcaaaag 2280
gccaggaacc gtaaaaaagg cgcgttgctg ctcaagttag aggtggcgaa acccgacagg actataaaga 2400
gagcatcaca aaaatcgacg ctcaagttag aggtggcgaa acccgacagg actataaaga 2400
taccaggcgt ttccccctgg tctcccttcg gtgcgctctc ggtgcgctctc cctgcccgtc 2460
accggatacc tgtccgcctt gttaggtcgt cgccttatcc actggttaaca acggtctacac tagctcacgc 2520
tgtaggtatc tcagttcggg ggcacgaacc gcctttatcc gctgtgagtc caaccggta 2640
ccggttcagc ccgaccgctg tatcgccact ctacagagtt cttgaagtgg gctgaagcca ggtggttttt 2700
agacacgact tctgcgctct aacaaaccac tcaagaagat ttaagggatt ttggtcatga tctaaagta 2760
gtaggcgggt ctacagagtt cttgaagtgg gctgaagcca ggtggttttt cctttgatct gattatcaaa 2820
gtatttggta tctgcgctct aacaaaccac tcaagaagat ttaagggatt ttggtcatga tctaaagta 2880
tgatccggca acggttcggg aaaaaggatc aaaactcacg ttttaaatga atgttaatc agtgaggcac 2940
acgcgagaa aaaaaggatc aaaactcacg ttttaaatga atgttaatc agtgaggcac 2940
cagtggaaag acctagatcc acttggtctg ccatagtgtg tgcaatgata agcgcgagac 3000
ttaccatctg gccccagtg cactcagttc cgggtcccaa ctccctcggt tatggcgaca 3060
ttatcagcaa taccagttat aagcggttag cactcagttc tttctgtgac cccggcgctc 3120
tccgcctcca tccagttat aagcggttag cactcagttc tttctgtgac cccggcgctc 3180
aatatggctt cactcagttc aagcggttag cactcagttc tttctgtgac cccggcgctc 3240
ggatggctt cactcagttc aagcggttag cactcagttc tttctgtgac cccggcgctc 3300
gtaagatgct cccggcgctc aagcggttag cactcagttc tttctgtgac cccggcgctc 3360
cggcgagcga actttaaaag gatccagttc ccagcggttc cgacacggaa atggtgaata 3420
actttaaaag gatccagttc ccagcggttc cgacacggaa atggtgaata 3480
ccgctgttga tttactttca ccagcggttc cgacacggaa atggtgaata 3540
tttactttca ccagcggttc cgacacggaa atggtgaata 3600
ggaataaggg agcattttat aaacaaatag ggggtccggc ctttattacc tgggtgagca 3660
agcattttat aaacaaatag ggggtccggc ctttattacc tgggtgagca 3720
aaacaaatag ggggtccggc ctttattacc tgggtgagca 3780
aaaaagcccc ccgaagcggg ggaagggcga tgcaaggcga ggggtgagca 3840
gcaactgttg ggggatgtgc gtaaaacgac tgataagata cattgatgag 3900
gtaaaacgac ggggatgtgc gtaaaacgac tgataagata cattgatgag 3960
tatacagaca aaaaaatgct ttatttggta aatttggat gctcagcatg 4020
aaaaaagcgt ggggtccggc ctttattacc tgggtgagca 4080
tgcaataaac aagttggggg gggcgagaa aaacgattcc cgtgtcagtc 4140
catccagcgg ggaatcgaaa tctcgtagca cgtgtcagtc 4140

```

<210> 118
 <211> 17384
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pFK161 Plasmid

<400> 118

gcgcacgagg	gagcttccag	ggggaaacgc	ctggtatctt	tatagtccctg	tcggggtttc	60
gccacctctg	acttgagcgt	cgatttttgt	gatgctcgtc	aggggggagg	agcctatgga	120
aaaacgccag	caacgcggcc	tttttacggg	tcctggccct	ttgctggcct	tttgctcaca	180
tggtctttcc	tgcggttatcc	cctgattctg	tggataaaccg	tattaccgcc	tttgagttag	240
ctgataccgc	tcgcccgcagc	cgaacgaccg	agcgcagcga	gtcagttagc	gaggaagcgg	300
aagagcgctg	acttccgcgt	ttccagactt	tacgaaacac	ggaaaccgaa	gaccattcat	360
gttggttgctc	aggtcgcaga	cgttttgtag	cagcagtcgc	ttcacgttcg	ctcgcgatc	420
gggtgattcat	tctgctaacc	agtaaggcaa	ccccgccagc	ctagccgggt	cctcaacgac	480
aggagcacga	tcatgcgcac	ccgtcagatc	cagacatgat	aagatacatt	gatgagtttg	540
gacaaaccac	aactagaatg	cagtgaaaaa	aatgctttat	ttgtgaaatt	tgtgatgcta	600
ttgcttttatt	tgtaaccatt	ataagctgca	ataaacaagt	taacaacaac	aattgcattc	660
atthttatggt	tcaggttcag	ggggagggtg	gggagggttt	ttaaagcaag	taaaacctct	720
acaaatgtgg	tatggctgat	tatgatctct	agtcaaggca	ctatacatca	aattattcctt	780
attaaccctt	ttacaaatta	aaaagctaaa	ggtacacaat	ttttgagcat	agttattaat	840
agcagacact	ctatgcctgt	gtggagttaag	aaaaaacagt	atgttatgat	tataactggt	900
atgcctactt	ataaagggtta	cagaatattt	ttccataatt	ttcttgtata	gcagtgcagc	960
tttttctctt	gtggtgtaaa	tagcaaagca	agcaagagtt	ctattactaa	acacagcatg	1020
actcaaaaaa	ctgaaggaaa	ctgaaggaaa	gtccttgagg	tcctctacct	ttctcttctt	1080
ttttggaggga	gtagaatggt	gagagtcagc	agtagcctca	tcatacactag	atggcatttc	1140
ttctgagcaa	aacaggtttt	cctcattaaa	ggcattccac	cactgctccc	attcatcagt	1200
tcctatagggt	ggaagtctaa	atacacaaac	aattagaatc	agtagtttaa	cacattatac	1260
acttaaaaaa	tttatattta	ccttagagct	ttaaatctct	gtaggtagtt	tgtccaatta	1320
tgtcacacca	cagaagtaag	gttccctcac	aaagatccgg	accaaagcgg	ccatcgtagc	1380
tcctccactcc	tgtagttcgg	gggcatggat	gcgcggatag	ccgctgctgg	tttccctggat	1440
gccgacggat	ttgcactgcc	ggtagaactc	gcgaggtcgt	ccagcctcag	gcagcagctg	1500
aaccaactcg	cgaggggatc	gagcccgagg	tgggcgaaga	actccagcat	gagatccccg	1560
cgctggaggga	tcacccagcc	ggcgtcccg	aaaacgattc	cgaagcccaa	cctttcatag	1620
aaggcgccgg	tggaatcgaa	atctcgtgat	ggcaggttgg	gcgtcgcttg	gtcgggtcatt	1680
tcgaacccca	gagtcgccgt	cagaagaact	cgtaagaag	gcgatagaag	gcgatgcgct	1740
gcgaatccgg	agcggcgata	cgtaaaagca	cgaggaagcg	gtcagcccat	tcgcccgaac	1800
gctcttcagc	aatatcacgg	gtagccaacg	ctatgtcctg	atagcgggtc	gccacaccca	1860
gccggccaca	gtcgatgaat	ccagaaaagc	ggccatttcc	caccatgata	ttcggcaagc	1920
aggcatcgcc	atgggtcacg	acgagatcct	cgccgtcggg	atgcgcgcct	tgagcctggc	1980
gaacagttcg	actggcgcca	gcccttgatg	ctcttcgtcc	agatcatcct	gatcgacaag	2040
accggcttcc	atccgagtag	gtgctcgtcc	gatgcgatgt	ttcgcttggt	ggtcgaatgg	2100
gcaggtagcc	ggatcaagcg	tatgcagccg	ccgcattgca	tcagccatga	tggtactttt	2160
ctcgccagga	gcaaggtgag	gcaaggtgag	atccgtcccc	ggcacttcgc	ccaatagcag	2220
ccagtcctct	cccgtttcag	tgacaacgtc	gagcacagct	gcgcaaggaa	cgcccgtcgt	2280
ggccagccac	gatagccgcg	ctgcctcgtc	ctgcagttca	ttcagggcac	cggacaggtc	2340
gggtcttgaca	aaaagaaccg	ggcgccctcg	cgctgacagc	cggaacacgg	cggcatcaga	2400
gcagccgatt	gtctgtttgt	cccagtcata	gccgaatagc	ctctccaccc	aagcgccggc	2460
agaacctgcg	tgcaatccat	cttggttcaat	cattgcgaac	gatcctcatc	ctgtctcttg	2520
atcagatctt	gatccctcgt	gccatcagat	ccttgccggc	aagaaagcca	ttcagttttac	2580
tttgacgggg	ttcccaacct	taccagaggg	cgccccagct	ggcaattccg	gttcgcttgc	2640
tgctccataaa	accgcccagt	ctagctatcg	ccatgtaagc	ccactgcaag	ctacctgctt	2700
tctcttttgcg	ccttgttcca	gatagcccag	gatagcccag	tagctgacat	tcacccgggg	2760
tcagcacccgt	ttctgcccag	tggttttcta	cgtgttccgc	ttccttttagc	agcccttgcg	2820
ccctgagtg	ttgcccgcag	gtgaaagctt	tttgcaaaa	cctaggcctc	caaaaaagcc	2880
tcctcactac	ttctggaata	gctcagaggg	cgagggcgcc	taataaaaaa	aaattagttc	2940
gccatggggc	ggagaatggg	cggaactggg	cggagttagg	ggcggggtgg	gcgaggttag	3000
gggcccggact	atgggttgctg	actaattgag	atgcattgct	tgcatacttc	tgccctgctg	3060
ggagcctggg	gactttccac	acctgggtgc	tgactaattg	agatgcatgc	tttgcatact	3120
tctgcctgct	ggggagcctg	gggactttcc	acaccctaac	tgacacacat	tcacagcccg	3180
gatctgcagg	acccaacgct	gcccagagatg	cgccgcgtgc	ggctgctgga	gatggcggag	3240
gcgatggata	tggtctgcca	aggggtgggt	tgccgattca	cagttctccg	caagaattga	3300
ttggctccaa	ttcttgaggt	ggtgaatccg	ttagcgaggt	gccgcgggct	tcattccagg	3360
tcgaggtggc	ccggctccat	gcaccgcgac	gcaacgcggg	gaggcagaca	aggtataggg	3420
cggcgcctac	aatccatgcc	aacccgttcc	atgtgctcgc	cgaggcgcat	aaatcgccgt	3480
gacgatcagc	ggtccaatga	tcgaagttag	gctggtaaga	gccgcgagcg	atccttgaag	3540
ctgtccctga	tggtcgtcat	ctacctgctt	ggacagcatg	gcctgcaacg	cggcatcccc	3600
atgcgcgcgg	aagcgagaag	aatcataatg	gggaaggcca	tcagccctcg	cgctcgcaac	3660
gccagcaaga	cgtagcccag	cgcgctcggg	cgccatgccc	gcgataatgg	cctgcttctc	3720
gccgaaacgt	ttggtggcgg	gaccagtgac	gaaggcttga	gcgagggcgt	gcaagattcc	3780
gaataaccga	agcgacagcg	cgatcatcgt	cgcgctccag	cgaaagcggt	cctcgccgaa	3840
aatgacccag	agcgctgccc	gcacctgtcc	tacgagttgc	atgataaaga	agacagtcac	3900
aagtgcggcg	acgatagtca	tgccccgcgc	ccaccgggaag	gagctgactg	ggttgaaggc	3960
tctcaagggc	atcgggtcgac	gctctccctt	atgcgactcc	tgcattagga	agcagccag	4020

tagtaggttg	aggccgttga	gcaccgccc	cgcaaggaat	ggtagcatgca	aggagatggc	4080
gcccacaggt	ccccgggcca	cgggcctgcc	accataccca	cgccgaaaca	agcgtcatg	4140
agcccgaggt	ggcgagcccg	atcttcccc	tcggtgatgt	cgccgatata	ggcgccagca	4200
accgcacctg	tggcgccggt	gatgccggcc	acgatgcgtc	cgccgtagag	gatcttggca	4260
gtcacagcat	gcgcataatcc	atgcttcgac	catgcgctca	caaagtaggt	gaatggcgca	4320
tgtagtaccc	acatcgtcat	cgctttccac	tgctctcgcg	aataaagatg	gaaaatcaat	4380
ctcatggtaa	tagtccatga	aaatccttgt	attcataaat	cctccaggta	gctatatgca	4440
aattgaaaca	aaagagatgg	tgatctttct	aagagatgat	ggaatctccc	ttcagtatcc	4500
cgatgggtcaa	tgcgctggat	atgggataga	tgggaatatg	ctgattttta	tgggacagag	4560
ttgcgaactg	ttcccaacta	aaatcatttt	gcacgatcag	cgcaactacga	actttaccca	4620
caaatagtca	ggtaatgaat	cctgatataa	agacaggttg	ataaatcagt	cttctacgcg	4680
catcgcacgc	gcacaccgta	gaaagtcttt	cagttgtgag	cctgggcaaa	ccgttaactt	4740
tcggcggtct	tgctgtgcga	caggctcacg	tctaaaagga	aataaatcat	gggtcataaa	4800
attatcacgt	tgctcggcgc	ggcgacggat	gttctgtatg	cgctgttttt	ccgtggcgcg	4860
ttgctgtctg	gtgatctgcc	ttctaaatct	ggcacagccg	aattgcgcga	gcttgggttt	4920
gctgaaacca	gacacacagc	aaactgaatac	cagaaagaaa	atcactttac	ctttctgaca	4980
tcagaagggc	agaaatattgc	cggtgaacac	ctgggtcaata	cgctgttttg	tgagcagcaa	5040
tattgcgctt	cgatgacgct	tggcggtgag	ctgctgcaca	aaaggcaatc	aaaggcaatc	5100
gacgagctgg	accagcgcat	tgctgacacc	gtctccttcg	aacttattcg	caatggagtg	5160
tcattcatca	aggacgcggc	tatcgcaaat	gggtgctatcc	acgcagcggc	aatcgaaaca	5220
cctcagccgg	tgaccaatat	ctacaacatc	agccttggtg	tccagcggtg	tgagccagcg	5280
cagaacaagg	taaccgctag	tggcgataag	ttcaaagtta	aacctgggtg	tgataccaac	5340
attgaaacgt	tgatcgaaaa	cgcgctgaaa	aacgctgctg	aatgtgcggc	gctggatgtc	5400
acaaagcaaa	tggcagcaga	caagaaagcg	atggatgaac	tggcttccta	tgcccgacag	5460
gccatcatga	tgggaatgttt	ccccgggtgg	gttatctggc	agcagtgccg	tcgatagtat	5520
gcaattgata	attattatca	tttgcggttc	ctttccggcg	atccgccttg	ttacggggcg	5580
gcgacctcgc	gggtttttcgc	tattttatgaa	aattttccgg	tttaaggcgt	ttcggttctt	5640
cttcgtcata	acttaatggt	tttattttaa	ataccctctg	aaaagaaaag	aaacgcagag	5700
tgctgaaagc	gagctttttg	gcctctgtcg	tttcctttct	ctgtttttgt	ccgtgggaatg	5760
aacaatggaa	gtcaacaaaa	agcagctggc	tgacattttc	ggtagcagta	ccgtaccat	5820
tcagaactgg	caggaaacagg	gaatgcccg	tctgcaggc	ggtagcagag	gtaatgaggt	5880
gctttatgac	tctgccgcgg	tcataaaaatg	gtatgccgaa	agggatgctg	aaattgagaa	5940
cgaaaagctg	cgccgggagg	ttgaagaact	gcggcagggc	agcgaggcag	atccacagga	6000
cggtgtgtgt	cgccatgatc	gcgtagtcca	tagtggctcc	aagttagcgaa	gcgagcagga	6060
ctgggcggcg	gcaaagcggt	cggacagtgc	tccgagaacg	ggtagcgcata	gaaattgcat	6120
caacgcata	agcgctagca	gcacgccata	gtgactggcg	atgctgtcgg	aatggacgat	6180
atccgcgaag	aggcccgcca	gtaccggcat	aaccaagcct	atgcctacag	catccagggt	6240
gacggtgccg	aggatgacga	tgagcgcatt	gttagatttc	atacacgggtg	cctgactgcg	6300
ttagcaattt	aactgtgata	aactaccgca	ttaaagctta	tcgatgataa	gcggtcaaac	6360
atgagaattc	gcggccgctc	ttctcgttct	gccagcgggc	cctcgtctct	ccaccccatc	6420
cgtctgccgg	tgggtgtgtg	aaggcagggg	tgcggctctc	cgccccgacg	ctgcccccg	6480
gcacttttct	tcagtgtgtc	gcgtgttctc	tgtaggtgtg	tgagggcgcc	gggtgtgtcc	6540
tcacgtgttt	cactttgttc	gtgtctcgct	tgaccattgt	cccagagtcg	gtggatgtgg	6600
ccggtggcgt	tgcataccct	tcccgtctgg	tgtgtgcacg	cgctgtttct	tgtaagcgct	6660
gaggtgctcc	tggagcgttc	caggtttgtc	tcctaggtgc	ctgcttctga	gctgggtggg	6720
tgctctgagaa	ttccctgggt	tgccctccgg	getccgtctg	gctgtgtgcc	ttcccgtttg	6780
tcgtcgggtg	gcccgtgaga	gggggggtcg	ggagagaagg	aggggcaaga	cccccttct	6840
ggtagcggtg	aggcgccac	cccgcgacta	gtacgcctgt	gcgtagggct	gggtgctgag	6900
gagaggcctg	ggggttggaa	agtttctcga	gagactcatt	gctttcccg	ggggagcttt	6960
aatgcccttg	gctttcgggg	gggaccggtt	gcagggtctc	ccctgtccgc	ggatgctcag	7020
ttgggtcttct	gaagagaacc	ttcctgttgc	cgcagacccc	cccgccgggt	cgcccgcgctg	7080
gtcgggggtt	gggttccctg	tgtgtctcgt	gcattgcatt	tctctcggtg	gcccgggctc	7140
ggaaagggtg	tgggtccgtc	ccgcctctcag	tgagaaagtt	tccttctcta	gctatcttcc	7200
ctcgccccct	cggtctctct	acggtctcga	gggtgtctct	ccgaatggct	ccctggaggg	7260
ggcctccccg	gacctctcc	cgcgcgcgca	gggtttgctc	tctcgtctac	cgccggcccg	7320
ccgttgctgc	gacccctcc	ggggagggat	cacgcggggc	agagcctgtc	tgtcgtctct	7380
gcacaccccc	ctccagattc	gctcgggttg	tgtggtgggt	ggctggggag	agggctccgt	7440
tgggtaggcg	ggagcatgtg	actttctctc	cctcctgagg	ggcgcggtgc	ggacgggggtg	7500
tcgctcgcgt	gcgtgcgcgt	cccggttccc	caccgctctt	cccgctgcctc	acccgtgcct	7560
ctgcgcgcgc	acggtggggt	cgctcgcgtc	cacgactttg	gcccgtcccg	cgacggcgccg	7620
cccccttcc	gcgtccctct	gctgtgtgct	tctcgggctg	tgtggttggt	tcgctctcgc	7680
cctcgggggt	cgtggtgctg	ttcccacggc	tggcgaaatc	gcgggagtc	tccttccctc	7740
gttctgtggg	cgccgcagcg	gtgtctggcg	ttggtctgag	tcgctctcgg	ggacggggacc	7800
gcccgtcggg	gagaggggtc	ttggcccgag	ccggcccgac	gtcggacgtg	gggacccact	7860
cggtgtcgcc	agaacggctg	ggtaggcac	gggtgtgctg	catcgggtctc	tctctcgtgt	7920
gggtgtggtg	tcctcggggt	cccggggggc	cgtcgtgttt	cggtcgggt	cgccgctgca	7980
	gactgctcag	gggagtggtg	cagtggtgatt	cccgcgggtt	ttgcctcgcg	8040

tgccttgacc	gggtccgacgc	ccgagcgggtc	tctcggtccc	ttgtgaggac	ccccctccgg	8100
gaggggccc	tttcggccgc	ccttgccgtc	gtcgccggcc	ctcggtctgc	tggtgcgttc	8160
ccccctccc	gctcgccgca	gccgggtctt	tttccctct	ccccccctct	cctctgactg	8220
acccgtggcc	gtgctgtcgg	accccccgca	tggggggcg	cgggcaagta	cgcgtccggg	8280
cggtcacccg	gggtcttggg	gggggcccag	gggtaagaaa	gtcgggtcgg	cgggcccggg	8340
gagctgtggt	ttggaggggc	lccccggccc	gcccgcgtgg	cgggtgtctt	cgcgggtctt	8400
gagagggctg	cgtgcgaggg	gaaaagggtt	ccccgcgagg	gcaaagggaa	agaggctagc	8460
agtggtcatt	gtcccgcacg	tggtggtggt	tggtggccga	ggtgcgtctg	gggggctcgt	8520
ccggccctgt	cgtccgtcgg	gaaggcgcgt	gttggggcct	gcccggagtgc	cgagggtggg	8580
accctggcgg	tgggattaac	cccgcgcgcg	tgctccgggt	tggcgggtgg	ggctccgggt	8640
gatgtctaac	tcctctccc	cgagggtctca	ggccttctcc	gcgcgggctc	tcggccctcc	8700
cctcggttct	ccctctcggc	gggttcaagt	cgtcgtcga	cctccccctc	tcggtccttc	8760
catctctcgc	gcaattggcg	cgcgcgagtt	cacgggtggg	tcgtcctccg	cctccgcttc	8820
tcgcgggggt	tgccgcctgt	tcgggtctct	cctgccgcac	ccccgttggc	gtggtctctt	8880
ctcgccgggt	tcgcggactc	ctggcttcgc	ccggagggtc	aggggggctt	ccggttcccc	8940
gacgttgcgc	ctcgtctcgt	tggtgttggg	ggggggccgc	tcgggcctcc	gcccgcctgc	9000
gagccccctg	cgcacccccc	ggtgtgcggg	ttcgccgcgc	ggtcagttgg	gcctggcggt	9060
tggtgcgcgt	cgggagcgtg	tcgcctcgc	ggcggttaga	cgcgggtgtc	gcccgggtcc	9120
gacgggtggc	ctatccaggg	ctcgccccgc	ccgacccccg	cctgcccgtc	ccgggtgggtg	9180
tcgttgggtg	ggggagtga	tggtgtctac	ggctcattcc	tcggcggtgg	tttgactgtc	9240
tcgcgggtgt	cgcgcttctc	tttcggccaa	ccccacgcgc	aacccaccc	cctgctctcc	9300
cggccccggt	cgggtcgacgt	tcgggtctct	ccgtagccga	gggggttcggg	atttgtgccc	9360
gggacggagg	ggagagcggg	taagagaggt	gtcggagagc	tgctccgggg	cgacgctcgg	9420
gttggctttg	ccgcgtcgtg	gtgctcgggg	acgggttttg	tcggaccccc	acgggggtcgg	9480
tcgggcgcga	tgactctctc	cgttcgcgcg	gagcgcgcgc	ccgggtcacc	cccggtttgt	9540
cctcccgcca	ggctctccgc	cgcgcgcgcg	tcctcctcct	ctctcgcgct	ctctgtcccc	9600
cctggctcct	tcacaccccc	gacgtccgcg	tcgcgttctc	ttacctgggt	gatcctgcca	9660
ggtagcatat	gcttgtctca	aagattaage	catgcatgtc	taagtacgca	cggccgggtac	9720
agtgaactg	cgaatggctc	attaaatcag	ttatgggtcc	tttgggtcgt	cgtcctctct	9780
ctacttggat	aactgtggta	attctagagc	taatacatgc	cgcagggcgc	tgacccccct	9840
tcgggggggg	ggatgcgtgc	atttatcaga	tcaaaaccaa	cccggtgagc	tcctcccggt	9900
ctccggccgg	gggtcggggc	ccggcggtct	ggtgactcta	gataacctcg	ggccgatcgc	9960
acgccccccc	tgggggcgac	gacccattcg	aacgtctgcc	ctatcaactt	tcgatggtag	10020
tcgcggtgcc	taccatgggt	accacgggtg	acgggggaat	aggggtcgat	tcgggagagg	10080
gagcctgaga	aacggctacc	acatccaagg	aaggcagcag	gcgcgcaaat	tacccactcc	10140
cgcacccggg	aggtagtgc	gaaaaataac	aatacaggac	tccttcgagg	ccctgtaatt	10200
ggaatgagtc	cactttaaat	cctttaacga	ggatccattg	gagggcaagt	ctgggtgccag	10260
cagccgcggg	aattccagct	ccaatagcgt	atattaaagt	tgctgcagtt	aaaaagctcg	10320
tagttggatc	ttgggagcgg	gcggggcggt	cgcgcgcagg	caggtcaccg	cccgcccccg	10380
ccccttgcc	ctcggcgcgc	cctcgatgct	cttagctgag	tgctccgcgg	ggcccgaagc	10440
gtttactttg	aaaaaattag	agtgttcaaa	gcaggcccca	gcccgcctgg	taccgcagct	10500
aggaataatg	gaatagacc	gcgggtctat	tttgttgggt	tcgggaactg	aggccatgat	10560
taagagggac	ggccgggggg	attcgtattg	cgcgcgctga	ggtgaaattc	ttggaccggc	10620
gcaagacgga	ccagagcgaa	agcatttgcc	aagaatgttt	tcattaatca	agaacgaaag	10680
tcggagggtc	gaagacgata	agataccgtc	gtagttccga	ccataaacga	tgccgatggg	10740
cgatgcggcg	gcgttattcc	catgaccgcg	cgggcagctt	ccgggaaacc	aaagtctttg	10800
gggtccgggg	ggagtatggt	tgcaaaagct	aaacttaag	gaattgacgg	aagggcacca	10860
ccaggagtgg	gcctgcggct	taatttgact	caacacggga	aacctcacc	ggcccgagca	10920
cggacaggat	tgacagattg	atagctcttt	ctcgattccg	tgggtgggtg	tgcatggccg	10980
ttcttagttg	gtggagcgat	ttgtctgggt	aattccgata	acgaacgaga	ctctggcatg	11040
ctaactagtt	acgcgacccc	cgagcggtcg	gcgtccccc	acttcttaga	gggacaagtg	11100
gcgttcagcc	acccgagatt	gagcaataac	aggtctgtga	tgcccttaga	tgtccggggc	11160
tgacgcgcg	ctacactgac	tggtccagcg	tggtccctacc	ctgcgcggcg	agggcggggt	11220
aaccggttga	accccattcg	tgatggggat	cggggattgc	aattattccc	catgaacgag	11280
gaattcccag	taagtgcggg	tcataagctt	gcgttgatta	agtccctgcc	ctttgtacac	11340
acgcgcgcgt	gctactaccg	attggatggt	ttagtggggc	cctcggtatg	gcccgcggcg	11400
ggtcggccca	cggccctggc	ggagcgtgga	gaagacgggt	gaacctgact	atctagagga	11460
agtaaaaagtc	gtaacaagggt	ttcgttaggt	gaacctgcgg	aaggatcatt	aaacggggaga	11520
ctgtggagga	cgggcggcgt	ggccgcgtct	ccccgtcttg	tggtgtgctc	cgcggggagg	11580
cgcgtgcgtc	cggggtcccg	tcgcccgcgt	tgtagcgcag	gtgtctggag	tgaggtgaga	11640
gaaggggtgg	gtgggggtcg	tctgggtccg	tctgggaccg	cctccgattt	cccccccccc	11700
tcctctctcc	ctcgtccggc	tctgacctgc	ccaccttacc	gcccgcggcg	ctgctcgcgg	11760
gcgtcttgc	tctttcccg	cggctcttcc	cgtgtctacg	aggggcgggt	cgtcgttacg	11820
ggtttttgac	cgcgtccggg	ggcggttcgg	cgtggggggc	gcgcgttttg	tctcccgcca	11880
cccatcccg	cgcgggtct	ggcttttcta	cgttgggtgg	ggcggttggt	gcgtgtgggg	11940
ggatgtgagt	gtcgcgtgtg	ggctcgcgcg	tcgcgatgcc	acgcttttct	ggcctcgcgt	12000
gtcctcccc	ctcgtgtccc	gggtacctag	ctgtcgcgtt	ccggcgcgga	ggtttaaggga	12060

ccccgggggg	gtcgccttgc	cgccccccag	gtcggggggg	gggtggggccc	gtaggggaagt	12120
cggttcgttc	ggcggtcttc	cctcagactc	catgaacctc	ctccccccgc	tgcgcctgtt	12180
cccagggcgg	cggttcgttg	gggggggtga	tgtctggagc	ccccctcggg	gccgtggggg	12240
cccagaccgc	gcccgcgggt	tgcccgatth	ccgcgggtcg	gtcctgtcgg	tgccgggtcgt	12300
gggttcccgt	gtcgttcccc	tggttttccg	ctccccgacc	tttttttttc	ctccccccca	12360
cacgtgtctc	gtttcgttcc	tgctggccgg	cctgaggcta	ccccctcggtc	catctgttct	12420
cctctctctc	cggggagagg	agggcggtgg	tcgttggggg	actgtgccgt	cgtcagcacc	12480
cgtcggttgc	ctcacacccg	aaataccgat	acgactctta	gcgggtggatc	actcggctcg	12540
tgcttcgatg	aagaacgcag	ctagctgcga	gaattaatgt	gaattgcagg	acacattgat	12600
catcgacact	tcgaacgcac	ttgcggcccc	gggttctctc	cggggctacg	cctgtcttag	12660
cgtcgggtga	cgatcaatcg	cgtcacccgc	tgcggtgggt	gctgcgcggc	tgggagtttg	12720
ctcgcagggc	caacccccca	acccgggtcg	ggccctccgt	ctcccgaagt	tcagacgtgt	12780
gggcggttgt	cggtgtggcg	cgcgcgcccg	cgtcgcggag	cctgggtctcc	cccgcgcatc	12840
cgcgctcgcg	gcttcgttcc	gctcgcgcgt	tcccgccttc	gcccgtgcac	cccggtcctg	12900
gcctcgcgtc	ggcgccctcc	ggaccgctgc	ctcaccagtc	tttctcggtc	ccgtgccccg	12960
tgggaaccca	cccgcccccc	gtggcgcccg	gggggtggcg	cgtccgcac	tgctctgggtc	13020
gagggtggcg	gttgagggtg	tgctgcgcgc	gaggtgggtg	tcggtccctt	gcggcgcccg	13080
gggtgtcggg	gtggcggtcg	acgagggccg	gtcggtcgcc	tgcggtgggt	gtctgtgtgt	13140
gtttgggtct	tgcgctgggg	gaggcggggt	cgaccgctcg	cggggttggc	gcgggtcgccc	13200
ggcgccgcgc	accctccggc	ttgtgtggag	ggagagcgag	ggcgagaacg	gagagaggtg	13260
gtatcccccg	tgccgttgcg	agggaggggt	tgccgtcccg	cgtccgtccg	tccctccctc	13320
cctcggtggg	cgcccttcgc	ccgcacgcgg	ccgctagggg	cggtcggggc	ccgtggcccc	13380
cgtgggtctt	cttcgtctcc	gcttctcctt	caccggggcg	gtaccgcgtc	cggcgcggc	13440
ccgcgggacg	ccgcggcgct	cgtgcgcga	tgcgagtcac	ccccgggtgt	tgcgagttcg	13500
gggagggaga	gggcccctcg	gacccgttgc	gtcccggctt	ccctgggggg	gacccggcgt	13560
ctgtgggctg	tgccgtcccg	gggttgccgt	tgagttaagt	cctccacccc	cgcccgcccg	13620
ccctcccgcc	ggccctctcg	ggacccctcg	agacgggtcg	ccggctcgtc	ctcccggtgc	13680
gcccgggtg	gtctctttcc	cgcccgccct	ctcgtctctt	tcttcccgcg	gctgggcgcg	13740
tgccccccct	ttctgaccgc	gacctagat	cagacgtggc	gacccgctga	atttaagcat	13800
attagtccag	ggaggaaaag	aaactaacca	ggattccctc	agtaacggcg	agtgaacagg	13860
gaagagccca	gcgcgcgaat	cccgccgcgc	gtcgcggcgt	gggaaatgtg	gcgtacggaa	13920
gacccactcc	ccggcgccgc	tcgtgggggg	cccaagtcct	tctgacgcag	gcccagcccg	13980
tggacgggtc	gagggccggta	gcggcccccg	cgcgcggggc	tcgggtcttc	ccggagtcgg	14040
gttgcttggg	aatgcagccc	aaagcggggt	gtaaactcca	tctaaggcta	aataccggca	14100
cgagaccgat	agtcaacaag	taccgtaagg	gaaagttgaa	aagaactttg	aagagagagt	14160
tcaagagggc	gtgaaaccgt	taagaggtaa	acgggtgggg	tccgcgcagt	ccgcggcgag	14220
gattcaaccc	ggcggcgcg	gtccggccgt	gcccgggtgt	ccccggcgat	ctttcccgtc	14280
ccccgttcc	cccagaccct	ccacccgcgc	gtcgttcccc	tcttctctcc	cgcgctccgg	14340
gcctccggcg	gcgggcgcgg	gggggtgggt	gggtgtggcg	cgcgggcggg	gcggggcggtg	14400
gggtcggcgg	gggaccgccc	ccggcccgcg	accggccgcc	gcccggcgca	cttccaccgt	14460
ggcggtgcgc	cgcgaccggc	tccgggacgg	ccgggaaggc	ccgggtggga	aggtggctcg	14520
ggggggggcg	cgctctcag	ggcgcccgca	accactcac	cccgagtgtt	acagccctcc	14580
ggcgcgcgtt	tcgcgcgaat	ccggggccga	ggaagccaga	taccgcctcg	cgcgctctcc	14640
ctctcccccc	gtccgcctcc	cgggcgggcg	tgggggtggg	ggccggggcg	ccccctccac	14700
ggcgcgaccg	ctctccacc	ccccctcgct	gcctctctcg	gggcccgggtg	ggggggcgggg	14760
cggactgtcc	ccagtgcgcc	ccgggcgtcg	tcgcgcctcg	gggtcccggg	gggaccgtcg	14820
gtcacgcgtc	tcccgcagaa	gcccgcgcga	cggggtcggc	ggcgatgtcg	gctaccacc	14880
cgacccgtct	tgaaacacgg	accaaagggt	ctaaccgcgt	cgcgagtcag	gggtcgtccc	14940
gaaagccgcc	gtggcgcaat	gaaggtgaag	ggccccgcgc	ggggggccga	ggtgggatcc	15000
cgaggcctct	ccagtccgcc	gagggcgcac	caccggcccg	tctcgcccgc	cgcgccgggg	15060
aggtggagca	cgagcgtacg	cgtaggacc	cgaaagatgg	tgaactatgc	ttgggcaggg	15120
cgaagccaga	ggaaactctg	gtggaggtcc	gtagcgggtc	tgacgtgcaa	atcggtcgtc	15180
cgacctgggt	ataggggcga	aagactaatc	gaaccatcta	gtagctgggt	ccctccgaag	15240
tttccctcag	gatagctggc	gctctcgtc	ccgacgtacg	cagttttatc	cggtaaagcg	15300
aatgattaga	ggtcttgggg	ccgaaacgat	ctcaacctat	tctcaaaact	taaattgggtg	15360
agaagcccg	ctcgtggcg	tggagccggg	cgtggaatgc	gagtgcctag	tgggcccactt	15420
ttggtaagca	gaactggcgc	tgccggatga	accgaacgcc	gggttaaggc	gcccgatgcc	15480
gacgctcatc	agaccccaga	aaaggtgttg	gttgatatag	acagcaggac	ggtggccatg	15540
gaagtgggaa	tccgctaagg	agtgtgtaac	aactcacctg	ccgaatcaac	tagccctgaa	15600
aatggatggc	gctggtagct	cgggcccata	cccggccgct	gcccagtcg	gaacggacgc	15660
ggacggggag	ggccgcgaat	tcttgaagac	gaaagggcct	cgtgatacgc	ctattttttat	15720
aggttaatgt	catgataata	atgggtttctt	agacgtcagg	tggcactttt	cggggaaatg	15780
tgccgggaac	ccctatttgt	ttattttttg	aaatatgtat	aaatatgtat	ccgctcatga	15840
gacaataacc	ctgataaatg	cttcaataat	attgaaaaag	gaagagtatg	agtattcaac	15900
atctccgtgt	cgcccttatt	cccttttttg	cggtcatttg	cttcctgttt	ttgctcacc	15960
agaaacgcgt	gtgaaagtaa	aagatgctga	agatcatttg	ggtgcacgag	tggtttacat	16020
cgaactggat	ctcaacagcg	gtaagatcct	tgagagtttt	cgccccgaag	aacgtttttc	16080

```

aatgatgagc actttttaaa tttctgctatg tggcgcggtg ttatcccggtg ttgacgcggg 16140
gcaagagcaa ctccggtcgcc gcatacacta ttctcagaat gacttggttg agtactcacc 16200
agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca gtgctgccat 16260
aaccatgagt gataaacttg cggccaactt acttctgaca acgatcggag gaccgaagga 16320
gctaaccgct tttttgcaca acatggggga tcatgtaact cgccttgatc gttgggaacc 16380
ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg cagcaatggc 16440
aacaacgttg cgcaactat taactggcga actacttact ctgcttccc ggcaacaatt 16500
aatagactgg atggaggcgg ataaaagttgc aggaccactt ctgcgctcgg cccttccggc 16560
tggtcggttt attgctgata aatctggagc cgggtgagcgt ggggtctcgg gtatcattgc 16620
agcactgggg ccagatggta agccctcccc tatcgtagtt atctacacga cggggagtc 16680
ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac tgattaagca 16740
ttggttaactg tcagaccaag tttactcata tatacttttag attgatttaa aacttcattt 16800
tcaattttaa aggatctagg tgaagatcct ttttgataat ctcatgacca aaactccctta 16860
acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag gatcttcttg 16920
agatcctttt tttctgcgcg taatctgctg cttgcaacaa aaaaaaccac cgctaccagc 16980
gggtggtttg ttgcccggatc aagagctacc aactcttttt ccgaaggtaa ctggcttcag 17040
cagagcgcag atacaaata ctgtccttct catacctcgc agtgtagccg tagttaggcc accacttcaa 17100
gaactctgta gcaccgccta tctgctaate ctgttaccag ctgttaccag 17160
cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac cggataaggg 17220
gcagcggctg ggctgaacgg ggggttctgtg cacacagccc agcttgagc gaacgaacta 17280
caccgaactg agatacctac agcgtgagct atgagaaagc gccacgcttc cgaagggaga 17340
aaggcggaca ggtatccggt aagcggcagg gtcggaacag gaga 17384

```

<210> 119
 <211> 2814
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pLITMUS38 Plasmid

```

<400> 119
gttaactacg tcaggtggca cttttcgggg aaatgtgcgc ggaaccacct tttgtttatt 60
tttctaaata cattcaaata tgtatccgct catgagacaa taacctgat aaatgcttca 120
ataatattga aaaaggaaga gtatgagtat tcaacatttc cgtgtcgccc ttattccctt 180
ttttcggcca ttttgccttc ctgtttttgc tcacccagaa acgctggtga aagtaaaaga 240
tgctgaagat cagttgggtg cactgagtggt ttacatcgaa ctggatctca acagcggtaa 300
gatccttgag agttttcgcc ccgaagaacg ttctccaatg atgagcactt ttaaagtctt 360
gctatgtggc gcggtattat cccgtgttga cgcggggcaa gagcaactcg gtcgcgcgat 420
acactattct cagaatgact tgggtgagta ctcaccagtc acagaaaagc atcttacgga 480
tggtcatgaca gtaagagaat tatgcagtcg tgccataacc atgagtata acactcgccc 540
caacttactt ctgacaacga tccgaggcta accgcttttt tgcacaacat 600
gggggatcat gtaactcgcc ttgatcgttg ggaaccggag ctgaatgaag ccataccaaa 660
cgacgagcgt gacaccacga cttactctag cttcccgga aactatgaag aactattaac 720
tggtcgaacta cttactctag gctcggccct ctcgcggtat acaattataa ggcgggataa 780
agttgcagga ccacttctgc gagcgtgggt acacgacggg gatcaggca actatggatg aacgaaatag 960
tggtgcccgt gtagttatct gagataggtg cctcactgat taagcattgg taactgtcag accaagttta 1020
ctcatatata ctttagattg atttaccctg gttgataatc agaaaagccc caaaaacagg 1080
aagattgtat aagcaaatat ttaaattgta aacgttaata ttttgttaaa attcgcgtta 1140
aatttttggt aaatcagctc attttttaac caataggccg aaatcggcaa aatcccttat 1200
aatcaaaagc aatagcccgga gatagggttg agtgtgttgc cagtttgtaa caagagtcca 1260
ctattaaaga acgtggactc caacgtcaaa cggcgaaaac ccgtctatca gggcgatggc 1320
ccactacgtg aaccatcacc caaatcaagt tttttggggg cgaggtgccg taaagcacta 1380
aatcggaacc ctaaaggagc ccccgatttt agagcttgac ggggaaagcg aacgtggcga 1440
gaaaggaaag gaagaaagcg aaaggagcgg ccgctagggc gctggcaagt gtagcgttca 1500
cgctgcgcgt aaccaccaca cccgcccgcg ttaatgcgcc gctacagggc gcgtaaaagg 1560
atctaggtga agatcctttt tgataatctc atgacaaaaa tcccttaacg tgagttttcg 1620
ttccactgag cgtacagacc cgtagaaaaa atcaaggatg cttcttgaga tccttttttt 1680
ctgcgcgtaa tctgtctgct gcaaacaaaa aaaccaccgc taccagcggg ggtttgtttg 1740
ccggtacaag agctaccaac tctttttccg aaggttaact gcttcagcag agcgcagata 1800
ccaaatactg ttcttctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca 1860
ccgctacat acctcgtctg gctaactctg ttaccagtgg ctgctgccag tggcgataag 1920
tcgtgtctta ccgggttgga ctcaagacga tagttaccgg ataaggcgca gcggtcgggc 1980
tgaacggggg gttcgtgcac acagccagc cgacctacac cgaactgaga 2040
tacctacgac gtgagctatg agaaagcgcc acgcttcccg aagggagaaa ggcggacagg 2100

```

tatccggtaa	gccccagggt	cggaacagga	gagcgacga	gggagcttcc	agggggaaac	2160
gcctggatc	tttatagtc	tgctgggtt	cgccacctct	gacttgagcg	tcgatttttg	2220
tgatgctcg	cagggggg	gagcctatgg	aaaaacgcca	gcaacgcggc	ctttttacgg	2280
ttcctggcct	tttgctggcc	ttttgctcac	atgtaatgtg	agttagctca	ctcattaggg	2340
accccaggct	ttacacttta	tgcttccggc	tcgtatgttg	tggtggaattg	tgagcgggata	2400
acaatttcac	acaggaaaca	gctatgacca	tgattacgcc	aagctacgta	atacgactca	2460
ctagtggggc	ccgtgcaatt	gaagccggct	ggcgccaagc	ttctctgcag	gatatctgga	2520
tcacgaatt	cgctagcttc	ggccgtgacg	cgtctccgga	tgtagaggca	tgcgctcgacc	2580
ctctagtcaa	ggccttaagt	gagtcgtatt	acggactggc	cgctcgtttta	caacgctcgtg	2640
actgggaaaa	ccctggcggt	acccaactta	atcgccctgc	agcacatccc	cctttcgcca	2700
gctggcgtaa	tagcgaagag	gcccgcaccg	atcgcccttc	ccaacagttg	cgcgccctga	2760
atggcgaatg	gcgcttcgct	tggttaataaa	gccccgttcg	gcgggctttt	tttt	2814

<210> 120

<211> 2847

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attB Plasmid

<400> 120

gttaactacg	tcaggtggca	cttttcgggg	aaatgtgcgc	ggaaccccta	tttgtttatt	60
tttctaaata	cattcaata	tgtatccgct	catgagacaa	taaccctgat	aaatgcttca	120
ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttattccctt	180
ttttgcgcca	ttttgccttc	ctgttttttg	tcaccagaaa	acgctgggtga	aagtaaaaga	240
tgctgaagat	cagttgggtg	cacgagtggg	ttacatcgaa	ctggatctca	acagcggtaa	300
gatccttgag	agttttcgcc	ccgaagaacg	ttctccaatg	atgagcactt	ttaaagttct	360
gctatgtggc	gcggtattat	cccggtgtga	cgccggggcaa	gagcaactcg	gtcgccgcat	420
acactattct	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	480
tggcatgaca	gtaagagaat	tatgcagtgc	tgccataacc	atgagtgata	acactgcggc	540
caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	tgcaacaacat	600
gggggatcat	gtaactcgcc	ttgatcgttg	ggaaccggag	ctgaatgaag	ccataccaaa	660
cgacgagcgt	gacaccacga	tgctgttagc	aatggcaaca	acgttgccga	aactattaac	720
tggcgaacta	cttactctag	cttcccggca	acaattaata	gactggatgg	aggcgggataa	780
agttgcagga	ccactctctg	gctcgccctt	tcgggtggc	tggtttattg	ctgataaaatc	840
tggagccggt	gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	atggtaagcc	900
ctcccgatc	gtagtattct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	960
acagatcgct	gagatagggt	cctcactgat	taagcattgg	taactgtcag	accaagttta	1020
ctcatatata	ctttagattg	atttaccctg	gttgataatc	agaaaagccc	caaaaacagg	1080
aagattgtat	aagcaaatat	ttaaattgta	aacgttaata	ttttgttaaa	attcgcgtta	1140
aatttttggt	aatatcagtc	attttttaac	caataggccg	aaatcgccaa	aatcccttat	1200
aatcaaaaag	aatagcccga	gatagggttg	agtgttggtc	cagtttgga	caagagtcca	1260
ctattaaaga	acgtggactc	caacgtcaaa	gggcgaaaaa	ccgtctatca	gggcgtaggc	1320
ccactacgtg	aaccatcacc	caaatcaagt	tttttggggt	cgaggtgccg	taaaagcata	1380
aatcggaacc	ctaaagggag	cccccgattt	agagcttgac	ggggaaagcg	aacgtggcga	1440
gaaaggaagg	gaagaaaagc	aaaggagcgg	gcgctagggc	gctggcaagt	gtagcgggtca	1500
cgctgcgcgt	aaccaccaca	cccgcgcgcg	ttaatgcgcc	gctacagggc	gcgtaaaagg	1560
atctaggtga	agatcctttt	tgataatctc	atgacccaaa	tcctttaacg	tgagttttcg	1620
ttccactgag	cgtcagaccc	cgtagaaaaa	atcaaaggat	cttcttgaga	tccttttttt	1680
ctgcgcgtaa	tctgctgctt	gcaaacaaaa	aaaccaccgc	taccagcggg	ggtttggttg	1740
ccggatcaag	agctaccaac	tctttttccg	aaggtaactg	gcttcagcag	agcgcagata	1800
ccaaatactg	ttcttctagt	gtagccgtag	ttaggccacc	acttcaagaa	ctctgtagca	1860
ccgcctacat	acctcgctct	gctaactctg	ttaccagtgg	ctgctgccag	tggcgataag	1920
tcgtgtctta	ccgggttgga	ctcaagacga	tagttaccgg	ataaggcgca	gcggtcgggc	1980
tgaacggggg	gttcgtgcac	acagcccagc	ttggagcgaa	cgacctacac	cgaactgaga	2040
tacctacagc	gtgagctatg	agaaaagcgc	acgcttcccg	aagggaagaaa	ggcggacagg	2100
tatccggtaa	gccccagggt	cggaacagga	gagcgacga	gggagcttcc	agggggaaac	2160
gcctggatc	tttatagtc	tgctgggtt	cgccacctct	gacttgagcg	tcgatttttg	2220
tgatgctcg	cagggggg	gagcctatgg	aaaaacgcca	gcaacgcggc	ctttttacgg	2280
ttcctggcct	tttgctggcc	ttttgctcac	atgtaatgtg	agttagctca	ctcattaggg	2340
accccaggct	ttacacttta	tgcttccggc	tcgtatgttg	tggtggaattg	tgagcgggata	2400
acaatttcac	acaggaaaca	gctatgacca	tgattacgcc	aagctacgta	atacgactca	2460
ctagtggggc	ccgtgcaatt	gaagccggct	ggcgccaagc	ttctctgcag	gattgaagcc	2520
tgctttttta	tactaaactg	agcgaaatct	ggatccacga	attcgctagc	ttcgccgctg	2580
acgcgtctcc	ggatgcgtacg	gcattgcgtc	accctctagt	caaggcctta	agtgagtcgt	2640
attacggact	ggccgctcgtt	ttacaacgtc	gtgactggga	aaaccctggc	gttacccaac	2700

ttaatgcgct	tgcagcacat	ccccctttcg	ccagctggcg	taatagcgaa	gaggcccgca	2760
ccgatcgccc	ttcccaacag	ttgcgcagcc	tgaatggcga	atggcgcttc	gcttggtaat	2820
aaagcccgct	tcggcgggct	ttttttt				2847

<210> 121
<211> 4223
<212> DNA
<213> Artificial Sequence

<220>
<223> pLIT38attBBSRpolyA2 Plasmid

<400> 121						
accatgaaaa	catttaacat	ttctcaacaa	gatctagaat	tagtagaagt	agcgacagag	60
aagattacaa	tgctttatga	ggataataaa	catcatgtgg	gagcggcaat	tcgtacgaaa	120
acaggagaaa	tcatttcggc	agtacatatt	gaagcgtata	taggacgagt	aactgtttgt	180
gcagaagcca	ttgcgattgg	tagtgacggt	tcgaatggac	aaaaggattt	tgacacgatt	240
gtagctgtta	gacaccctta	ttctgacgaa	gtagatagaa	gtattcgagt	ggtaagtcct	300
tgtggtatgt	gtaggggagt	gatttcagac	tatgcaccag	attgttttgt	gttaatagaa	360
atgaatggca	agttagtcaa	aactacgatt	gaagaactca	ttccactcaa	atatacccga	420
aattaaaaagt	ttaccatac	caagcttggc	tgctgcctga	ggctggacga	cctcgcggag	480
ttctacccgg	agtgcacatc	cgtcggcatc	caggaaaacca	gcagcggcta	tccgcgcatac	540
catgcccccg	aactgcagga	gtggggaggc	acgatggccg	ctttgggtccg	gatctttgtg	600
aaggaaacctt	acttctgtgg	tgtgacataa	ttggacaaaac	tacctaagaa	gatttaaagc	660
tctaagggtaa	atataaaaatt	tttaagtgtg	taatgtgtta	aactactgat	tctaatttgt	720
tgtgtatttt	agattccaac	ctatggaact	gatgaatggg	agcagtgggt	gaatgccttt	780
aatgaggaaa	acctgttttg	ctcagaagaa	atgccatcta	gtgatgatga	ggctactgct	840
gactctcaac	attctactcc	tccaaaaaag	aagagaaaag	tagaagaccc	caaggacttt	900
ccttcagaat	tgctaagttt	tttgagtcac	gctgtgttta	gtaatagaac	tcttgcttgc	960
tttgcatttt	acaccacaaa	ggaaaaagct	gaactgctat	acaagaaaaa	tatggaaaaa	1020
tattctgttaa	cctttataag	taggcataac	agttataatc	ataacatact	gttttttctt	1080
actccacaca	ggcatagagt	gtctgtctatt	aataactatg	ctcaaaaatt	gtgtaccttt	1140
agctttataat	ggttacaagt	ggtttaataag	gaatatttga	tgtatagtgc	cttgactaga	1200
gatcataatc	agccatacca	catttgtaga	ggttttactt	gctttaaaaa	acctcccaca	1260
cctccccctg	aacctgaaac	ataaaaatgaa	tgcaatttgt	gttgtaact	tgtttattgc	1320
agcttataat	ggttacaaat	aaagcaatag	catcacaaat	ttcacaaaata	aagatccaga	1380
tttcgctcaa	gttagtataa	aaaagcaggc	ttcaatcctg	cagagaagct	tggcgccagc	1440
cggttccaat	tgacggggcc	ccactagtga	gtcgtattac	gtagcttggc	gtaatcatgg	1500
tcatagctgt	ttcctgtgtg	aaattgttat	ccgctcacaa	ttccacacaa	catacgagcc	1560
ggaagcataa	agtgtaaagc	ctgggggtgcc	taatgagtga	gctaactcac	attacatgtg	1620
agcaaaaagg	cagcaaaaagg	ccaggaaccg	taaaaaggcc	gcgttgctgg	cgttttttcca	1680
tttgctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	tcaagtccaga	gggtggcgaaa	1740
cccagacagga	ctataaagat	accaggcggt	tccccctgga	agctccctcg	tgcgctctcc	1800
tgttccgacc	ctgcccgtta	ccggataact	gtccgccttt	ctcccttcgg	gaagcgtggc	1860
gctttctcat	agctcacgct	gtaggtatct	cagttcgggtg	taggtcgttc	gctccaagct	1920
gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	gtaactatcg	1980
tcttgagtc	aacccggtaa	gacacgactt	atcgccactg	gcagcagcca	ctggtaacag	2040
gattagcaga	gcgaggtatg	taggcggtgc	tacagagttc	ttgaagtggg	ggcctaacta	2100
cggctacact	agaagaacag	tatttggtat	ctgcgctctg	ctgaagccag	ttaccttcgg	2160
aaaaagagtt	ggtagctctt	gatccggcaa	acaaaaccacc	gctggtagcg	gtggtttttt	2220
tgtttgcaag	cagcagatta	cgcgacagaa	aaaaggatct	caagaagatc	ctttgatctt	2280
ttctacgggg	tctgacgctc	agtggaaacga	aaactcacgt	taagggattt	tgggtcatgag	2340
attatcaaaa	aggatcttca	cctagatcct	tttacgcgcc	ctgtagcggc	gcattaaagc	2400
cggcgggtgt	ggtgggttac	cgcagcgtga	ccgctacact	tgccagcgcc	ctagcgcgcc	2460
ctccctttcg	tttcttccct	tcctttctcg	ccacgttcgc	tttccccgct	aagctctaaa	2520
tcggggggct	cctttagggt	tccgatttag	tgctttacgg	cacctcgacc	ccaaaaaact	2580
tgatttgggt	gatgggtcac	gtagtgggct	atcgccctga	tagacgggtt	ttcgcccttt	2640
gacgttggag	tccacgttct	ttaatagtgg	actcttggtc	caaactggaa	caacactcaa	2700
ccctatctcg	ggctattctt	ttgatttata	agggattttg	ccgatttcgg	cctattgtgt	2760
aaaaaatgag	ctgatttaac	aaaaatttaa	cgcgaatttt	aacaaaatat	taacgtttac	2820
aatttaataa	tttgcttata	caatcttctt	gtttttgggg	cttttctgat	tatcaaccgg	2880
ggtaaatcaa	tctaaagtat	atatgagtaa	acttgggtctg	acagttacca	atgcttaacc	2940
agtgaaggac	ctatctcagc	gatctgtcta	tttcggtcat	ccatagttgc	ctgaactccc	3000
gtcgtgtaga	taactacgtg	acgggagggc	ttaccatctg	gccccagtg	tgcaatgata	3060
ccgcgagacc	cacgctcacc	ggctccagat	ttatcagcaa	taaaccagcc	agccgggaag	3120
gccgagcgca	gaagtgggtc	tgcaacttta	tccgcctcca	tccagtcct	taattgttgc	3180
cgggaagcta	gagtaagtag	ttcgccaggt	aatagtttgc	gcaacgttgt	tgccattgct	3240

acaggcatcg	tggtgtcacg	ctcgtcgctt	ggataggctt	cattcagctc	cgggtcccaa	3300
cgatcaaggc	gagttacatg	atcccccatg	ttgtgcaaaa	aagcgggttag	ctccttcggt	3360
cctccgatcg	ttgtcagaag	taagttggcc	gcagtggtat	cactcatggt	tatggcagca	3420
ctgcataatt	ctcttactgt	catgccatcc	gtaagatgct	tttctgtgac	tggtgagtac	3480
tcaaccaagt	cattctgaga	atagtgtatg	cggcgaccga	gttgctcttg	ccggcgctca	3540
acacgggata	ataccgcgcc	acatagcaga	actttaaaag	tgctcatcat	tggagaacgt	3600
tcttcggggc	gaaaactctc	aaggatctta	ccgctgttga	gatccagttc	gatgtaaccc	3660
actcgtgcac	ccaactgatc	ttcagcatct	tttactttca	ccagcgtttc	tggttgagca	3720
aaaacaggaa	ggcaaaaatgc	cgcaaaaaag	ggaataaggg	cgacacggaa	atgttgaata	3780
ctcatactct	tcttttttca	atattattga	agcatttatc	agggttattg	tctcatgagc	3840
ggatacatat	ttgaatgtat	ttagaaaaat	aaacaaatag	gggttcgcgc	cacatttccc	3900
cgaaaagtgc	cacctgacgt	agttaacaaa	aaaaagcccg	ccgaagcggg	ctttattacc	3960
aagcgaaagc	ccattcgcca	ttcaggctgc	gcaactgttg	ggaagggcga	tcggtgcggg	4020
cctcttcgct	attacgccag	ctggcgaaag	gggtagtgcc	tgcaaggcga	ttaaagtggg	4080
taacgccagg	gtttttcccag	tcacgacgtt	gtaaaacgac	ggccagtcgc	taatacgact	4140
cacttaaggc	cttgactaga	gggtcgacgc	atgcctgtac	atccggagac	gcgtcacggc	4200
cgaagctagc	gaattcgtgg	atc				4223

<210> 122

<211> 2686

<212> DNA

<213> Artificial Sequence

<220>

<223> pUC18 Plasmid

<400> 122

tcgcgcgctt	cggatgatgac	ggatgaaaacc	tctgacacat	gcagctcccc	gagacgggtca	60
cagcttgtct	gtaagcggat	gccggggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	120
ttggcgggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	180
accatatgcg	gtgtgaaata	ccgcacagat	gcgtaaaggag	aaaataccgc	atcaggcgcc	240
attcgccatt	caggctgcgc	aactgttggg	aagggcgatc	ggatggggcc	tcttcgctat	300
tacgccagct	ggcgaaaagg	ggatgtgctg	caaggcgatt	aagttgggta	acgccagggt	360
tttcccagtc	acgacgttgt	aaaacgacgg	ccagtgccta	gcttgcatgc	ctgcaggctc	420
actctagagg	atccccgggt	accgagctcg	aattcgtaat	catggtcata	gctgtttcct	480
gtgtgaaatt	gttatccgct	cacaattcca	cacaacatac	gagccggaag	cataaagtgt	540
aaagcctggg	gtgcctaata	agttagctaa	ctcacattaa	ttgcgttgcg	ctcactgccc	600
gctttccagt	cgggaaacct	gtcgtgccag	ctgcattaat	gaatcggcca	acgcgcgggg	660
agagggcggt	tgcgtattgg	gcgctcttcc	gcttccctgc	tcactgactc	gctgcgctcg	720
gtcgttcggc	tgcggcgagc	ggatcagct	cactcaaagg	cggtaatagc	gttatccaca	780
gaatcagggg	ataacgcagg	aaagaacatg	tgagcaaaag	gccagcaaaa	ggccagggaac	840
cgtaaaaagg	cgcgcttgct	ggcgttttcc	cataggtctc	gccccctga	cgagcatcac	900
aaaaatcgac	gctcaagtca	gaggtggcga	aaacccgacag	gactataaag	ataccaggcg	960
tttccccctg	gaagctccct	cgtgcgctct	cctgttcoga	ccctgccgct	taccggatac	1020
ctgtccgctt	ttctcccttc	gggaagcgtg	cgcttttctc	atagctcacg	ctgtaggtat	1080
ctcagttcgg	tgtaggctcg	tcgctccaag	ctgggctgtg	tgcacgaacc	ccccgttcag	1140
cccgaccgct	ggcgcttata	cggtaactat	cgtcttgagt	ccaacccggt	aagacacgac	1200
ttatcgccac	tggcagcagc	cactggtaac	aggattagca	gagcgaggta	tgtaggcggt	1260
gctacagagt	tcttgaagtg	gtggcctaac	tacggctaca	ctagaaggac	agtatttggt	1320
atctgcgctc	tgtatgaagc	agttaccttc	ggaaaaagag	ttggtagctc	ttgatccggc	1380
aaacaaacca	ccgctggtag	cgggtggttt	tttgtttgca	agcagcagat	tacgcgcaga	1440
aaaaaaggat	ctcaagaaga	tcttttgatc	ttttctacgg	ggcttgacgc	tcagtggaaac	1500
gaaaactcac	gttaagggat	tttggtcatg	agattatcaa	aaaggatctt	cacctagatc	1560
cttttaaaatt	aaaaatgaag	ttttaaatca	atctaaagta	tatatgagta	aacttgggtc	1620
gacagttacc	aatgcttaat	cagttaggca	cctatctcag	cgatctgtct	atttcgttca	1680
tccatagttg	cctgactccc	cgtcgtgtag	ataactacga	tacgggaggg	cttaccatct	1740
ggccccagtg	ctgcaatgat	accgcgagac	ccacgctcac	cggctccaga	tttatcagca	1800
ataaaccagc	cagccggaag	ggccgagcgc	agaagtgggtc	ctgcaacttt	atccgcctcc	1860
atccagctca	ttaatgtgtg	ccgggaagct	agagtaagta	gttcgcccag	taatagtttg	1920
cgcaacgctg	ttgccattgc	tacaggcatc	gtgggtgtcac	gctcgtcgtt	tggtatggct	1980
tcattcagtg	ccggttccca	acgatcaagg	cgagttacat	gatcccccat	gttgtgcaaa	2040
aaagcgggta	gctccttcgg	tcctccgcat	gttgtcagaa	gtaagtggc	cgcagtggtc	2100
tcactcatgg	ttatggcagc	actgcataat	tctcttactg	tcattgcatc	cgtaagatga	2160
ttttctgtga	ctgggtgagta	ctcaaccaag	tcattctgag	aatagtgat	gcccgcaccg	2220
agttgctctt	gcccggcgct	aatacggggg	aataccgcgc	cacatagcag	aactttaaaa	2280
gtgctcatca	ttgggaaaac	ttcttcgggg	cgaaaactct	caaggatctt	accgctgttg	2340
agatccagtt	cgatgtaacc	cactcgtgca	cccaactgat	cttcagcatc	ttttactttc	2400

accagcggtt	ctgggtgagc	aaaaacagga	aggcaaaatg	ccgcaaaaaa	gggaataagg	2460
gcgacacgga	aatgttgaat	actcatactc	ttcctttttc	aatattattg	aagcatttat	2520
cagggttatt	gtctcatgag	cggatacata	tttgaatgta	tttagaaaaa	taaacaaata	2580
ggggttccgc	gcacatttcc	ccgaaaagtg	ccacctgacg	tctaagaaac	cattattatc	2640
atgacattaa	cctataaaaa	taggcgtatc	acgaggccct	ttcgtc		2686

<210> 123

<211> 8521

<212> DNA

<213> Artificial Sequence

<220>

<223> pCXeGFPattB(6xHS4)2 Plasmid

<400> 123

tacggggcgg	gggatccact	agttattaat	agtaatcaat	tacgggggtca	ttagttcata	60
gcccataat	ggagtccgc	gttacataac	ttacggtaaa	tggcccgccct	ggctgaccgc	120
ccaacgaccc	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	acgccaatag	180
ggactttcca	ttgacgtcaa	tgggtggact	atttacggta	aactgcccac	ttggcagtag	240
atcaagtgtg	tcatatgcc	agtacgcccc	ctattgacgt	caatgacggg	aaatggcccc	300
cctggcatta	tgcccagtag	atgaccttat	gggactttcc	tacttggcag	tacatctacg	360
tattagtcac	cgctattacc	atgggtcgag	gtgagcccca	cgttctgctt	cactctcccc	420
atctcccccc	cctccccacc	cccaattttg	tatttattta	ttttttaatt	attttgtgca	480
gcgatggggg	cggggggggg	ggggggcgcg	gccaggcggg	gcggggcggg	gcgagggggc	540
gggcgggggc	aggcgagag	gtgcgggcgc	agccaatcag	agcggcgcg	tccgaaaggt	600
tccttttatg	gcgaggcggc	ggcgggcggc	gccctataaa	aagcgaagcg	cgcgggcggg	660
gggagtcgct	gcgttgccct	cgccccgtgc	cccgcctcgc	gccgcctcgc	gccggccggc	720
ccggctctga	ctgaccgcgt	tactcccaca	ggtagcgggg	cgggacggcc	cttctcctcc	780
gggctgtaat	tagcgcttgg	tttaatgacg	gctcgtttct	tttctgtggc	tgcgtgaaag	840
ccttaaaggg	ctccgggagg	gccctttgtg	cgggggggag	cggtcgggg	gggtcggtgc	900
tgtgtgtgtg	cgtggggagc	gccgcgtgcg	gccgcgcgtg	cccggcggtc	gtgagcgctg	960
cgggcgggcg	gcggggcctt	gtgcgctccg	cgtgtgcgcg	aggggagcgc	ggccgggggg	1020
ggtgccccgc	ggtagcgagg	ggtagcgagg	ggaaacaaag	ctgctgtgcg	gggtgtgctg	1080
tggggggggtg	agcagggggg	gtggggcgcg	cggtcgggct	gtaaccccc	cctgcacccc	1140
cctccccag	ttgctgagca	cgccccggct	tcgggtgcgg	ggctcgtg	ggggcggtgg	1200
gcggggcctg	ccgtgcccgg	cggggggtgg	cggggggtgg	gggtgcgggg	cgggggcggg	1260
ccgcctcggg	ccggggaggg	ctcgggggag	gggcgcggcg	gccccggagc	gccggcggtc	1320
gtcgaggcgc	ggcgagccgc	agccattgcc	ttttatggta	atcgtgcgag	agggcgagg	1380
gacttctctt	gtcccaaatc	tggcggagcc	gaaatctggg	aggcgccg	gcacccctc	1440
tagcggggcg	ggcggaagcg	gtgcccgcgc	ggcaggaagg	aaatgggcgg	ggagggcctt	1500
cgtgcgtcgc	cgccgcggcg	tccccttctc	catctccagg	ctcggggctg	ccgcaggggg	1560
acggctgcgc	tcggggggga	cgggggcagg	cggggttcgg	cttctggcgt	gtgacggcg	1620
gctctagagc	ctctgctaac	catgttcatg	ccttcttctt	tttctacag	ctcctgggca	1680
acgtgtggtg	tggtgtgctg	tctcatcatt	ttggcaaaag	attcgccacc	atggtagagca	1740
agggcgagga	gctgttcacc	gggggtgggc	ccatcctggt	cgagctggac	ggcgacgtaa	1800
acggccacaa	gttcagcgtg	tccggcgagg	gcgagggcga	tgccacctac	ggcaagctga	1860
ccctgaagtt	catctgcacc	accggcaagg	tgcccgtgcc	ctggcccacc	ctcgtgacca	1920
ccctgaccta	cggcgtgcag	tgcttcagcc	gctaccccga	ccacatgaag	cagcacgact	1980
tcttcaagtc	cgccatgccc	gaaggctacg	tccaggagcg	caccatcttc	ttcaaggacg	2040
acggcaacta	caagaccgcg	gccgaggtga	agttcgaggg	cgacaccctg	gtgaaccgca	2100
tcgagctgaa	gggcatcgac	ttcaaggagg	acggcaacat	cctggggcac	aagctggagt	2160
acaactacaa	cagccacaac	gtctatatca	tggccgacaa	gcagaagaac	ggcatcaagg	2220
tgaacttcaa	gatccgccac	aacatcgagg	acggcagcgt	gcagctcgcc	gaccactacc	2280
agcagaacac	ccccatcggc	gacggccccg	tgctgctgcc	cgacaaccac	tacctgagca	2340
cccagtcgcg	cctgagcaaa	gaccccaacg	agaagcgcg	tcacatgggtc	ctgctggagt	2400
tcgtgaccgc	cgccgggatc	actctcgga	tggacgagct	gtacaagtaa	gaattcactc	2460
ctcaggtgca	ggctgcctat	cagaaggtgg	tggtgtgtgt	ggccaatgcc	ctggctcaca	2520
aataccactg	agatcttttt	ccctctgcca	aaaattatgg	ggacatcatg	aagccocttg	2580
agcatctgac	ttctggctaa	taaaggaaat	ttattttcat	tgcaatagtg	tggttgaatt	2640
ttttgtgtct	ctcactcgga	aggacatatg	ggaggggcaa	tcatttaaaa	catcagaatg	2700
agtatttggg	ttagagtgtg	gcaacatatg	ccatattgctg	gctgccatga	acaaagggtg	2760
ctataaagag	gtcatcagta	tatgaaacag	ccccctgctg	tccattcctt	attccataga	2820
aaagccttga	cttgagggtta	gatttttttt	atatatttgtt	ttgtgttatt	ttttctctta	2880
acatccctaa	aattttctct	acatgtttta	ctagccagat	tttctctcct	ctcctgacta	2940
ctcccagtca	tagctgtccc	tcttctctta	tgaagatccc	tcgacctgca	gcccagctat	3000
ggcgtaatca	tggtcatagc	tgtttctctg	gtgaaattgt	tatccgctca	caattccaca	3060
caacatacga	gccggaagca	taaagtgtaa	agcctggggg	gcctaattgag	tgagtttaact	3120

cacattaatt	gcggttgcgct	cactgcccgc	tttccagtcg	ggaaacctgt	cgtgccagcg	3180
gatccgcac	tcaatttagtc	agcaaccata	gtcccgcccc	taactccgcc	catecccgcc	3240
ctaactccgc	ccagttccgc	ccattctccg	ccccatggct	gactaatttt	ttttatttat	3300
gcagagggcg	aggccgcctc	ggcctctgag	ctattccaga	agtagtgagg	aggctttttt	3360
ggaggctagt	ggatcccccg	ccccgtatcc	cccagggtgc	tgcaggctca	aagagcagcg	3420
agaagcgttc	agaggaaagc	gatcccgctc	caccttcccc	gtgcccgggc	tgtccccgca	3480
cgctgccggc	tcggggatgc	ggggggagcg	ccggaccgga	gcggagcccc	ggggcggtcg	3540
ctgctgcccc	ctagcggggg	agggacgtaa	ttacatccct	gggggctttg	ggggggggct	3600
gtccccgtga	gcggatccgc	ggccccgtat	cccccagggt	tctgcaggct	caaagagcag	3660
cgagaagcgt	tcagaggaaa	gcgatcccg	gccacgttcc	ccgtgcccgg	gctgtccccg	3720
cacgctgccg	gctcggggat	gcggggggag	cgccggaccg	gagcggagcc	ccggcggtcg	3780
cgctgctgcc	ccctagcggg	ggagggacgt	aattacatcc	ctgggggctt	tggggggggg	3840
ctgtccccgt	gagcggatcc	gcggccccgt	atcccccagg	tgtctgcagg	ctcaaagagc	3900
agcgagaaag	gttcagagga	aagcgatccc	gtgccacctt	ccccgtgccc	gggtgtccc	3960
cgcacgctgc	cggctcgggg	atgcgggggg	agcgccggac	cggagcggag	ccccggggcg	4020
ctcgtctgct	ccccctagcg	ggggagggac	gtaattacat	ccctgggggc	tttggggggg	4080
gctgtgtcccc	gtgacggat	ccgcggcccc	gtatccccca	ggtgtctgca	ggctcaaaga	4140
gcagcgagaa	gcgttcagag	gaaagcgatc	ccgtgccacc	ttccccgtgc	ccgggctgtc	4200
ccccgcacgt	gcgggctcgg	ggatgcgggg	ggagcgccgg	accggagcgg	agccccgggc	4260
ggctcgctgc	tgcctccctag	cgggggaggg	acgtaattac	atccctgggg	gctttggggg	4320
ggggctgtcc	ccgtgagcgg	atccgcggcc	ccgtatcccc	cagggtgtctg	caggctcaaa	4380
gagcagcgag	aagcgttcag	aggaaaagcga	tccccgtgcca	ccttccccgt	gccccggctg	4440
tccccgcacg	ctgcgggctc	ggggatgcgg	ggggagcgcc	ggaccggagc	ggagccccgg	4500
gcggctcgct	gctgccccct	agcgggggag	ggacgttaatt	acatccctgg	gggctttggg	4560
gggggggctgt	ccccgtgagc	ggatccgcgg	ccccgtatcc	cccagggtgc	tgcaggctca	4620
aagagcagcg	agaagcgctc	agaggaaaagc	gatcccgctc	caccttcccc	gtgcccgggc	4680
tgtccccgca	cgctgccggc	tcggggatgc	ggggggagcg	ccggaccgga	gcggagcccc	4740
gggcgggctcg	ctgctgcccc	ctagcggggg	agggacgtaa	ttacatccct	gggggctttg	4800
gggggggggct	gtccccgtga	gcggatccgc	ggggctgcag	gaattcgatt	gaagcctgct	4860
tttttatact	aaacttgagcg	aaatcaagct	cctaggcttt	tgcaaaaagc	taacttgttt	4920
attgcagctt	ataatgggta	caaataaagc	aatagcatca	caaatttcac	aaataaagca	4980
ttttttttcac	tgcatctetg	ttgtgggttg	tccaaactca	tcaatgtatc	ttatcatgtc	5040
tggatccgct	gcattaatga	atcgcccaac	gcgcggggag	aggcggtttg	cgtattgggc	5100
gctcttccgc	ttcctcgctc	actgactcgc	tgcgctcggt	cggttcggctg	cgccgagcgg	5160
tatcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	atcaggggat	aacgcaggaa	5220
agaacatgtg	agcaaaaagg	cagcaaaaag	ccaggaaccg	taaaaaggcc	gcgttgctgg	5280
cgttttttcca	taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	tcaagtcaag	5340
gggtggcgaaa	cccgacagga	ctataaagat	accaggcggt	tccccctgga	agctccctcg	5400
tgcgctctcc	tgttccgacc	ctgcgcgtta	ccggataacct	gtccgccttt	ctcccttggg	5460
gaagcgtggc	gctttctcaa	tgcctacgct	gtaggatatct	cagttcgggtg	taggtcggtc	5520
gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	5580
gtactatctg	tcttgagttc	aaccgcgtta	aacacgactg	atcgccactg	gcagcagcca	5640
ctggtaacag	gattagcaga	gcgaggtatg	taggcgggtc	tacagagttc	ttgaagtggg	5700
ggcctaacta	cggctacact	agaaggacag	tatttgggat	ctgcgctctg	ctgaagccag	5760
ttaccttccg	aaaaaggtt	ggtagctctt	gatccggcaa	acaaaaccacc	gctggtagcg	5820
gtgggttttt	tgtttgcaag	cagcagatta	cgccgagaaa	aaaaggatct	caagaagatc	5880
ctttgatctt	ttctacgggg	tctgacgtc	agtggaaacg	aaactcacgt	taagggatct	5940
tggatcatgag	attatcaaaa	aggatcttca	cctagatcct	tttaaattaa	aaatgaagtt	6000
ttaaatcaat	ctaaagtata	tatgagtaaa	cttgggtctga	cagttaccaa	tgtttaatca	6060
gtgaggcacc	tatctcagcg	atctgtctat	ttcggttcac	catagttgcc	tgactccccg	6120
tcgtgtagat	aactacgata	cgggagggct	taccatctgg	ccccagtgct	gcaatgatac	6180
cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	gccggaaggg	6240
ccgagcgcag	aagtgggtcct	gcaactttat	ccgcctccat	ccagttctatt	aattgttgcc	6300
gggaagctag	agtaagtagt	tcgccagtta	atagttttgc	caacgtttgt	gccattgcta	6360
caggcatcgt	ggtgtcacgc	tcgtcggttg	gtatggcttc	attcagctcc	ggttcccaac	6420
gatcaaggcg	agttacatga	tccccatgt	tgtgcaaaaa	agcggttagc	tccttcggtc	6480
ctccgatcgt	tgtcagaagt	aagtgtggcg	cagtgttatc	actcatgggt	atggcagcac	6540
tgcataattc	tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	ggtgagtact	6600
caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	ccggcgctca	6660
tacgggataa	taccgcgcca	catagcagaa	ctttaaaagt	gctcatcatt	ggaaaaagct	6720
cttcggggcg	aaaactctca	aggatcttac	cgctgttgag	atccagttcg	atgtaaccca	6780
ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	cagcgtttct	gggtgagcac	6840
aaacaggaag	gcaaaaagcc	gcaaaaaggg	gaataagggc	gacacggaaa	tgttgaaata	6900
tcatactctt	cctttttcaa	tattattgaa	gcatttatca	gggttattgt	ctcatgagcg	6960
gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	ggttccgcgc	acatttcccc	7020
gaaaagtgc	acctgggtcg	cggtatcgat	aagcttgata	tcgaattcct	gcagccccgc	7080
ggatccgctc	acgggggacag	ccccccccca	aagccccag	ggatgtaatt	acgtccctcc	7140

```

cccgcctaggg  ggcagcagcg  agccgccccg  ggctccgctc  cggctccggcg  ctccccccgc  7200
atccccgagc  cggcagcgtg  cgggggacagc  ccggggcacgg  ggaaggtggc  acgggatcgc  7260
tttcctctga  acgcttctcg  ctgctctttg  agcctgcaga  cacctggggg  atacgggggc  7320
gcggatccgc  tcacggggac  agcccccccc  caaagcccc  agggatgtaa  ttacgtccct  7380
ccccgcctag  ggggcagcag  cgagccgccc  ggggctccgc  tccgggtccg  cgctcccccc  7440
gcatccccga  gccggcagcg  tgcggggaca  gcccgggcac  ggggaaggtg  gcacgggatc  7500
gctttcctct  gaacgcttct  cgctgctctt  tgagcctgca  gacacctggg  ggatacgggg  7560
ccgcggatcc  gctcacgggg  acagcccccc  cccaaagccc  ccagggatgt  aattacgtcc  7620
ctcccccgct  agggggcagc  agcagccgc  ccgggggtcc  gctccgggtc  ggcgtcccc  7680
ccgcaccccc  gagccggcag  cgtgccccga  cagcccgggc  acggggaaag  tggcacggga  7740
tcgctttcct  ctgaacgctt  ctgctgctc  tttgagcctg  cagacacctg  ggggatacgg  7800
ggccgcggat  ccgctcacgg  ggacagcccc  cccccaaagc  cccagggat  gtaattacgt  7860
ccctcccccc  ctagggggca  gcagcgagcc  gcccggggct  ccgctccgg  ccggcgtcc  7920
ccccgcaccc  ccgagccggc  agcgtgcccc  gacagcccg  gcacggggaa  ggtggcacgg  7980
gatcgctttc  ctctgaacgc  ttctcgctgc  tctttgagcc  tgcagacacc  tgggggatac  8040
ggggccgcgg  atccgctcac  ggggacagcc  ccccccaaaa  gccccagg  atgtaattac  8100
gtccctcccc  cgctaggggg  cagcagcgag  ccgccccggg  ctccgctccg  gtcggcgct  8160
ccccccgcac  ccccgagccg  gcagcgtgcg  gggacagccc  gggcacgggg  aaggtggcac  8220
gggatcgctt  tcctctgaac  gcttctcgct  gctctttgag  cctgcagaca  cctgggggat  8280
acggggccgc  ggatccgctc  acggggacag  ccccccccca  aagccccag  ggatgtaatt  8340
acgtccctcc  cccgctaggg  ggcagcagcg  agccgccccg  ggctccgctc  cggctccggc  8400
ctccccccgc  atccccgagc  cggcagcgtg  ccggggacagc  ccggggcacg  ggaaggtggc  8460
acgggatcgc  tttcctctga  acgcttctcg  ctgctctttg  agcctgcaga  cacctggggg  8520
a  8521

```

<210> 124
 <211> 8851
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> p18EPOcDNA Plasmid

```

<400> 124
cagttgcccc  cggggctcgc  cagggcgaa  tccgccccc  acggctgctc  gccgatctcg  60
gtcatggccg  gccccgagcc  gtccccga  ttctgtggac  cgacctccga  ccactcggcg  120
tacagctcgt  ccaggcccg  caccacacc  cagggtcacg  tggtgtccgg  caccacctgg  180
tccctggacc  cgctgatgaa  cagggtcac  tccgtccgga  ccacaccggc  gaagtccgtc  240
tccacgaagt  cccgggagaa  cccgagccg  tccgtccaga  actcgaccgc  tccggcgacg  300
tcgcgcgcgg  tgagcaccgg  aacggcact  gtcaacttgg  ccatggatcc  agatttctcg  360
caagttagta  taaaaaagca  ggcttcaat  ctgcagagaa  gcttgatata  gaattcctgc  420
agcccccgcc  atccgctcac  ggggacagcc  ccccccaaaa  gccccagg  atgtaattac  480
gtccctcccc  cgctaggggg  cagcagcgag  ccgccccggg  ctccgctccg  gtccggcgct  540
ccccccgcac  ccccgagccg  gcagcgtg  gggacagccc  gggcacgggg  aaggtggcac  600
gggatcgctt  tcctctgaac  gcttctcgct  gctctttgag  cctgcagaca  cctgggggat  660
acggggccgc  ggatccgctc  acggggacag  ccccccccca  aagccccag  ggatgtaatt  720
acgtccctcc  cccgctaggg  ggcagcagcg  agccgccccg  ggctccgctc  cggctccggc  780
ctccccccgc  atccccgagc  cggcagcgtg  ccggggacagc  ccggggcacg  ggaaggtggc  840
acgggatcgc  tttcctctga  acgcttctcg  ctgctctttg  agcctgcaga  cacctggggg  900
atacggggcc  gcggatccgc  tcacggggac  agcccccccc  caaagcccc  agggatgtaa  960
ttacgtccct  cccccgctag  ggggcagcag  cgagccgccc  ggggctccgc  tccggtccgg  1020
cgctcccccc  gcatccccga  gccggcagcg  tgccggggac  gcccgggcac  ggggaaggtg  1080
gcacgggatc  gctttcctct  gaacgcttct  cgctgctctt  tgagcctgca  gacacctggg  1140
ggatacgggg  ccgcggtacc  gctcacgggg  acagcccccc  cccaaagccc  ccagggatgt  1200
aattacgtcc  ctcccccgct  agggggcagc  agcagccgc  ccgggggtcc  gctccgggtc  1260
ggcgctcccc  ccgcatcccc  gagccggcag  cgtgccccga  cagcccgggc  acggggaaag  1320
tggcacggga  tcgctttcct  ctgaacgctt  ctogctgctc  tttgagcctg  cagacacctg  1380
ggggatacgg  ggccgcggat  ccgctcacgg  ggacagcccc  cccccaaagc  cccagggat  1440
gtaattacgt  cctccccccg  ctagggggca  gcagcgagcc  gcccggggct  ccgctccgg  1500
ccggcgctcc  ccccgcatcc  ccgagccggc  agcgtgcccc  gacagcccg  gcacggggaa  1560
gggtggcacg  gatcgctttc  ctctgaacgc  ttctcgctgc  tctttgagcc  tgcagacacc  1620
tgggggatac  gggggccggg  atccgctcac  ggggacagcc  ccccccaaaa  gccccagg  1680
atgtaattac  cgctaggggg  cgtagccgag  ccgagccggg  ccgccccggg  ctccgctccg  1740
gtccggcgct  cccccgcac  ccccgagccg  gcagcgtg  gggacagccc  gggcacgggg  1800
aaggtggcac  ggatcgctt  tcctctgaac  gcttctcgct  gctctttgag  cctgcagaca  1860
cctgggggat  acggggcg  ggatccacta  gtaataaata  gtaataaata  acggggtcat  1920
tagttcatag  cccatatatg  gagttccgcg  ttacataact  tacggtaaat  ggcccgccct  1980

```

gctgaccgcc	caacgacccc	cgcccatgga	cgtcaataat	gacgtatgtt	cccatagtaa	2040
cgccaatagg	gactttccat	tgacgtcaat	gggtggacta	tttacggtaa	actgcccact	2100
tggcagtaca	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggcccg	ctggcattat	gcccagtaca	tgaccttatg	ggactttcct	acttggcagt	2220
acatctacgt	attagtcac	gctattacca	tgggtcgagg	tgagccccc	gttctgcttc	2280
actctcccca	tctccccccc	ctccccaccc	ccaattttgt	atthatttat	tttttaatta	2340
ttttgtgcag	cgatgggggc	gggggggggg	ggggcgcgcg	ccaggcgggg	cgggggcggg	2400
cgagggggcg	ggcggggcga	ggcggagagg	tggcgcgcca	gccaatcaga	gcggcgcgct	2460
ccgaaagttt	ccttttatgg	cgaggcgggc	gcggcgggcg	ccctataaaa	agcgaagcgc	2520
gcggcgggcg	ggagtgcgtg	cgttgccctc	gccccgtgcc	ccgctccgcg	ccgcctcgcg	2580
ccgccccccc	cggctctgac	tgaccgcgtt	actcccacag	gtgagcgggc	gggacggccc	2640
ttctcctccg	ggctgtaatt	agcgtttgg	ttaatgacgg	ctcgtttcct	ttctgtggct	2700
gcgtgaaagc	cttaaagggc	tccgggaggg	ccctttgtgc	ggggggggagc	ggctcggggg	2760
gtgcgtgcgt	gtgtgtgtgc	gtggggagcg	cccgctgcgg	cccgcgctgc	ccggcggtgc	2820
tgagcgctgc	ggcgcgggcg	cggggctttg	tgcgtccgc	gtgtgcgcga	ggggagcgcg	2880
gcccggggcg	gtgccccgcg	gtgccccggg	gctgcgaggg	gaacaaaggg	tgcggtcggg	2940
gtgtgtgcgt	gggggggtga	gcaggggtgt	tggcgcgccg	ggtcgggctg	taaccccccc	3000
ctgcaccccc	ctccccgagt	tgctgagcac	ggccccggctt	cggtgcgggg	gctccgtgcg	3060
gggcgtggcg	cggggctcgc	cgtgccccgc	gggggggtggc	ggcaggtggg	ggtgccccgg	3120
ggggcggggc	cgccctgggc	cggggagggc	tcgggggagg	ggcgcgggcg	ccccggagcg	3180
ccggcggtgc	tcgagggcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcgaga	3240
gggcgcaggg	acttcctttg	tcccaaatct	ggcggagccg	aaatctggga	ggcgcccgcg	3300
cacccccctc	agcgggcgcg	ggcgaagcgg	tcggcgcccg	gcaggaagga	aatggcgggg	3360
gagggccttc	gtgcgtcgcc	gcgcgcgcgt	ccccctctcc	atctccagcc	tcggggctgc	3420
cgagggggga	cggctgcctt	cggggggggac	ggggcagggc	ggggttcggc	ttctggcggt	3480
tgaccggcg	ctctagaatg	gggggtgcacg	aatgtcctgc	ctggctgtgg	cttctcctgt	3540
ccctgctgtc	gctccctctg	ggcctcccag	tcctggcgcg	cccaccacgc	ctcatctgtg	3600
acagccgagt	cctggagagg	tacctcttgg	aggccaagga	ggccgagaat	atcacgacgg	3660
gctgtgctga	acactgcagc	ttgaatgaga	atatcactgt	cccagacacc	aaagttaatt	3720
tctatgcctg	gaagaggatg	gaggtcgggc	agcaggccgt	agaagtctgg	cagggcctgg	3780
ccctgctgtc	ggaagctgtc	ctgcggggcc	aggccctgtt	ggtcaactct	tcccagccgt	3840
gggagccccc	gcagctgcac	gtggataaag	ccgtcagtgg	ccttcgcagc	ctcaccactc	3900
tgcttcgggc	tctgggagcc	cagaagggaag	ccatctcccc	tccagatgcg	gcctcagctg	3960
ctccactccg	aaacaatcact	gctgacactt	tccgcaaaact	cttccgagtc	tactccaatt	4020
tccctccggg	aaagctgaag	ctgtacacag	gggaggccctg	caggacaggg	gacagatgac	4080
gtacaagtaa	gaattcactc	ctcaggtgca	ggctgcctat	cagaaggtgg	tggtcggtgt	4140
ggccaatgcc	ctggctcaca	aataccactg	agatcttttt	ccctctgcca	aaaattatgg	4200
ggacatcatg	aagccccctg	agcatctgac	ttctggctaa	taaaggaaat	ttattttcat	4260
tgcaatcctt	ttttgtgtct	ttttgtgtct	ctcactcgga	aggacatatg	ggagggcaca	4320
tcatttaaaa	catcagaatg	agtatttgg	ttagagtgtg	gcaacatatg	ccatatgtctg	4380
gctgccatga	acaaaagggtg	ctataaagag	gtcatcagta	tatgaaacag	ccccctgtctg	4440
tccattcctt	attccataga	aaagccttga	cttgaggtta	gatttttttt	atattttgtt	4500
ttgtgttatt	ttttctttta	acatccctaa	aattttccctt	acatgtttta	ctagccagat	4560
ttttcctcct	ctcctgacta	ctcccagtc	tagctgtccc	tcttctctta	tgaagatccc	4620
tcgacctgca	gcccaggtt	gcattgcctgc	ctcactgcac	tagtggatcc	cccgcgccgt	4680
atcccccagg	tgtctgcagg	ctcaaagagc	agcgagaagc	gttcagagga	aagcgatccc	4740
gtgccacctt	ccccgtgccc	gggctgtccc	cgcacgctgc	cggtcgggg	atgccccggg	4800
agcgccggac	cggagcgagg	ccccggggcg	ctcgctgctg	ccccctagcg	ggggagggac	4860
gtaattacat	ccctgggggc	tttggggggg	ggctgtcccc	gtgagcggat	ccgcggcccc	4920
gtatccccca	gggtgtctga	ggctcaaaga	gcagcgagaa	gcgttcagag	gaaagcgatc	4980
ccgtgcaccc	ttccccgtgc	ccgggctgtc	cccgcacgct	gccggctcgg	ggatcggggg	5040
ggagcgccgg	accggagcgg	agccccgggc	ggctcgctgc	tgccccctag	cgggggaggg	5100
acgttaattac	atccctgggg	gctttggggg	ggggctgtcc	ccgtgagcgg	atccgcggcc	5160
ccgtatcccc	caggtgtctg	caggctcaaa	gagcagcgag	aagcgttcag	aggaaagcga	5220
tcccgctgcca	ccttccccgt	gcccgggctg	tccccgcacg	ctgccggctc	ggggatgcgg	5280
ggggagcgcc	ggaccggagc	ggagcccccg	gcccgtcgct	gctgccccct	agccccgggag	5340
ggacgttaatt	acatccctgg	gggctttggg	ggggggctgt	ccccgtgagc	ggatcccggg	5400
ccccgtatcc	cccaggtgtc	tgcaggctca	aagagcagcg	agaagcgttc	agaggaaagc	5460
gatccccgtc	caccttcccc	gtgccccggg	tgtccccgca	cgctgccccg	tcggggatgc	5520
ggggggagcg	cccgaccggg	gcggagcccc	ggcgccgtcg	ctgctgcccc	ctagcggggg	5580
agggacgttaa	ttacatccct	gggggctttg	ggggggggct	gtccccgtga	gcggatccgc	5640
ggccccgtat	cccccagggt	tctgcaggct	caaagagcag	cgagaagcgt	tcagaggaaa	5700
gcgatcccg	gccaccttcc	ccgtgccccg	gtgtccccg	cacgtgccc	gctcggggg	5760
gcgggggggag	cgccggaccg	gagcggagcc	ccggggcggt	cgctgctgccc	ccctagcggg	5820
ggagggacgt	aattacatcc	ctgggggctt	tggggggggg	ctgtccccgt	gagcggatcc	5880
gcggccccgt	atccccagg	tgtctgcagg	ctcaaagagc	agcgagaagc	gttcagagta	5940
aagcgatccc	gtgccacctt	ccccgtgccc	gggctgtccc	cgcacgctgc	cggtcgggg	6000

atgccccggggg	agcgccgggac	cgagagcgggag	ccccggggcg	ctcgctgctg	ccccctagcg	6060
ggggaggggac	gtaattacat	ccctggggggc	tttggggggg	ggctgtcccc	gtgagcggat	6120
ccgccccggct	gcaggaattc	gtaatcatgg	tcatagctgt	ttcctgtgtg	aaattgttat	6180
ccgctcacaa	ttccacacaa	catacgagcc	ggaagcataa	agtgtaaagc	ctgggggtgcc	6240
taatgagtga	gctaactcac	attaattgag	ttgcgctcac	tgccccgctt	ccagtcggga	6300
aacctgtcgt	gccagctgca	ttaatgaatc	ggccaacgag	cggggagagg	cggtttgcgt	6360
attggggcgct	cttccgcttc	ctcgctcact	gactcgctgc	gctcggtcgt	tgggctgagg	6420
cgagcggtat	caactcactc	aaaggcggtg	atacggttat	ccacagaatc	aggggataac	6480
gcaggaaaga	acatgtgagc	aaaaggccag	caaaaggcca	ggaaccgtaa	aaaggccgag	6540
ttgctggcggt	ttttccatag	gctccgcccc	cctgacgagc	atcacaaaaa	tgcagcgtca	6600
agtcagaggt	ggcgaaaacc	gacaggacta	taaagatacc	aggcgtttcc	ccctgggaagc	6660
tccctcgtgc	gctctcctgt	tccgaccctg	ccgcttaccg	gatacctgtc	cgcctttctc	6720
ccttcgggaa	gcgtggcgct	ttctcatagc	tcaagctgta	ggtatctcag	ttcggtgtag	6780
gtcgttcgct	ccaagctggg	ctgtgtgcac	gaaccccccg	ttcagcccca	ccgctgcggc	6840
ttatccggta	actatcgtct	tgagtccaac	ccggtaaagc	acgacttatt	gccactggca	6900
gcagccactg	gtaacaggat	tagcagagcg	aggatgttag	gcggtgctac	agagttcttg	6960
aagtcaggtt	ggcgaaaacc	ctacactaga	aggacagtat	ttggtatctg	cgctctgctg	7020
aagccagttt	ccttcggaaa	aagagtttgt	agctcttgat	ccggcaaaaa	aaccaccgct	7080
ggtagcgggt	gtttttttgt	ttgcaagcag	cagattacgc	gcagaaaaaa	aggatctcaa	7140
gaagatcgtt	tgatcttttc	tacggggtct	cagcgtcagt	ggaacgaaaa	ctcacgttaa	7200
gggatttttg	tcatgagatt	atcaaaaaag	atcttcacct	agatcctttt	aaattaaaaa	7260
tgaagtttta	aatcaatcta	aagtatatat	gagtaaaact	gggtctgacag	ttaccaatgc	7320
ttaatcagtg	aggcacctat	ctcagcgatc	tgtctatttc	gttcattccat	agttgcctga	7380
ctccccgctg	tgtagataac	tacgatacgg	gagggcttac	catctggccc	cagtgcgtgc	7440
atgataccgc	gagacccacg	ctcaccgggt	ccagatttat	cagcaataaa	ccagccagcc	7500
ggaagggcgg	agcgacagaag	tggtcctgca	actttatccg	cctccatcca	gtctattaat	7560
tggtgcgggg	aagctagagt	aagtagttcg	ccagtttaata	gtttgcgcaa	cgttggttgc	7620
attgctacag	gcacgtgggt	gtcacgctcg	tcgtttggtg	tggtctcatt	cagctccggt	7680
tcccaacgat	caaggcgagt	tacatgatcc	cccatggtgt	gcaaaaaagc	ggttagctcc	7740
ttcggtcctc	gggataatag	cagaagtaag	ttggccgcag	tggtatcact	catggttatg	7800
gcagcactgc	ataattctct	tactgtcatg	ccatccgtaa	gatgcttttc	tgtgactggg	7860
gagtactcaa	ccaagtcatt	ctgagaatag	tgtatgcggc	gaccgagttg	ctcttgcccc	7920
gcgtcaatac	cgccgacacat	cgccgcaact	agcagaactt	taaaagtgtc	catcattgga	7980
aaacgttctt	cggggcgaaa	actctcaagg	atcttaccgc	tggtgagatc	cagttcagatg	8040
taacccactc	gtgcacccaa	ctgatcttca	gcacttttta	ctttcaccag	cgtttctggg	8100
tgagcaaaaa	caggaaggca	aaatgcccga	aaaaagggaa	taagggcgac	acggaaatgt	8160
tgaataactc	tactcttctc	ttttcaatat	tattgaagca	tttatcaggg	ttattgtctc	8220
atgagcggtg	acatatttga	atgtattttag	aaaaataaac	aaataggggt	tccgcgcaca	8280
tttccccgaa	aagtgccacc	tgacgtagtt	aacaaaaaaa	agcccgcgca	agcgggcttt	8340
attaccaagc	gaagcgccat	tgcgccattc	ggctgcgcaa	ctggtgggaa	gggcgatcgg	8400
tgcgggcctc	ttcgctatta	cgccagctgg	cgaaaggggg	atgtgctgca	aggcgattaa	8460
ggtggaatac	gggataatag	tcccagtcac	gacgttgtta	aacgacggcc	agtcggtaat	8520
acgactcact	taaggccttg	actagagggg	cgacggtata	cagacatgat	aagatacatt	8580
gatgagtttg	gacaaaccac	aactagaatg	cagtgaaaaa	aatgctttat	ttgtgaaatt	8640
tgtgatgcta	ttgctttatt	tgtaaccatt	ataagctgca	ataaacaagt	tggggtgggc	8700
gaagaactcc	agcatgagat	ccccgcgctg	gaggatcatc	cagccggcgt	cccggaatac	8760
tattccgaag	cccaaccttt	catagaaggc	ggcggtggaa	tcgaaatctc	gtagcacgtg	8820
tcagtctctg	tctcgggcca	cgaagtgcac	g			8851

<210> 125

<211> 10474

<212> DNA

<213> Artificial Sequence

<220>

<223> p18genEPO Plasmid

<400> 125

cagttgcggg	ccgggtcgcg	cagggcgaac	tcccccccc	acggctgctc	gccgatctcg	60
gtcatggccg	gcccggagcg	gtcccgaag	ttcgtggaca	cgacctccga	ccactcggcg	120
tacagctcgt	ccaggccgag	cacccacacc	caggccaggg	tggtgtccgg	caccacctgg	180
tccctggacc	cgctgatgaa	cagggtcacg	tcgctccgga	ccacaccggc	gaagtgcgtc	240
tccacgaagt	cccgggagaa	cccgagccgg	tccgtccaga	actcgaccgc	tccggcgacg	300
tccgcgcggg	tgagcaccgg	aacggcactg	gtcaacttgg	ccatggatcc	agatttcgct	360
caagttagta	taaaaaagca	ggcttcaatc	ctgcagagaa	gcttgatata	gaattcctgc	420
agcccccgga	atccgctcac	ggggacagcc	ccccccaaa	gccccagggg	atgtaattac	480
gtccctcccc	cgctaggggg	cagcagcgag	ccgccccggg	ctccgctccg	gtccggcgct	540

ccccccgcat	ccccgagccg	gcagcgtgcg	gggacagccc	gggcacgggg	aaggtggcac	600
gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	cctgggggat	660
acggggccgc	ggatccgctc	acggggacag	cccccccca	aagccccag	ggatgtaatt	720
acgtccctcc	cccgcctagg	ggcagcagcg	agccgcccgg	ggctccgctc	cggctccggcg	780
ctccccccgc	atccccgagc	cggcagcgtg	cggggacagc	ccgggcacgg	ggaaggtggc	840
acgggatcgc	tttccctctga	acgtctctcg	ctgctctttg	agcctgcaga	cacctggggg	900
atacgggggc	gcggatccgc	tcacggggac	agcccccccc	caaagccccc	agggatgtaa	960
ttacgtccct	cccccgctag	ggggcagcag	cgagcgcgcc	ggggctccgc	tcgggtccgg	1020
cgctcccccc	gcatccccga	gccggcagcg	tgccggggaca	gcccggggcac	gggggaaggtg	1080
gcacggggatc	gcttttccctc	gaacgcttct	cgctgctctt	tgagcctgca	gacacctggg	1140
ggatacgggg	cccgcgatcc	gctcacgggg	acagcccccc	cccaaagccc	ccagggatgt	1200
aattacgtcc	ctcccccgct	agggggcagc	agcgagccgc	ccggggctcc	gctccgggtcc	1260
ggcgctcccc	ccgcatecccc	gagccggcag	cgtgcgggga	cagcccgggg	acgggggaagg	1320
tggcacggga	tcgcttttcc	ctgaacgctt	ctcgctgctc	tttgagcctg	cagacacctg	1380
ggggatacgg	ggccgcggat	ccgctcacgg	ggacagcccc	cccccaaagc	ccccagggat	1440
gtaattacgt	ccctcccccg	ctagggggca	gcagcgagcc	gcccgggggt	ccgctccggg	1500
ccggcgctcc	cccccgatcc	ccgagccggg	gttatttaata	gacagcccg	gcacggggaa	1560
gggtggcacgg	gatecgctttc	ctctgaacgc	ttctcgctgc	tccttgagcc	tgacagaccc	1620
tgggggatac	ggggccgcgg	atccgctcac	ggggcagagcc	ccccccaaa	gcccccaggg	1680
atgtaattac	gtccctcccc	cgctaggggg	cagcagcgag	ccgcccgggg	ctccgtcccg	1740
gtccggcgct	ccccccgcat	ccccgagccg	gcagcgtgcg	gggacagccc	gggcacgggg	1800
aaggtggcac	gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	1860
cctgggggat	acggggcggg	ggatccacta	gttatttaata	gtaatcaatt	acgggggtcat	1920
tagttcatag	cccatatag	gagttccgcg	ttacataact	tacggtaaat	ggcccgcctg	1980
gctgacccgc	caacgacccc	cgcccattga	cgtaataaat	gacgtatgtt	cccatagtaa	2040
cgccaatagg	gactttccat	tgactcaat	gggtggacta	tttacggtaa	actgccact	2100
tggcagtaca	tcaagtgat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggcccg	ctggcattat	gcccagtaaca	tgaccttatg	ggactttcct	acttggcagt	2220
acatctacgt	attagtcac	gctattacca	tgggtcgagg	tgagccccac	gttctgcttc	2280
actctcccc	ctcccccccc	ctcccccccc	ccaattttgt	atattattat	tttttaatta	2340
ttttgtgcag	cgatgggggg	gggggggggg	ggggcgcgcg	ccaggcgggg	cggggcgggg	2400
cgagggggcg	ggcggggcga	ggcggagagg	tgccggcgga	gccaatcaga	gcggcgcgct	2460
ccgaaagt	ccctttatgg	cgaggcgcg	gcggcgggcg	ccctataaaa	agcgaagcgc	2520
gcggcgggcg	ggagtcgctg	cgltgccttc	gccccgtgcc	ccgctccgcg	ccgcctccgc	2580
ccgcccgcgc	cggtcttgac	tgaccgcgtt	actcccacag	gtgagcgggc	gggacgggcc	2640
ttctctcccg	ggctgtaatt	agcgcttggt	ttaatgacgg	ctcgtttctt	ttctgtggct	2700
gcgtgaaagc	cttaaagggc	tcggggaggg	cccttttgtg	gggggggagc	ggctcggggg	2760
gtgcgtgcgt	gtgtgtgtgc	gtggggagcg	cccgctgcgg	cccgcgctgc	ccggcgggctg	2820
tgagcgtgc	gggcgcggcg	cggggctttg	tcgctccgc	gtgtgcgcga	ggggagcgcg	2880
gccggggcg	gtgccccgcg	gtgccccggg	gctgcgaggg	gaacaaagge	tgcggtgcgg	2940
gtgtgtgcgt	gggggggtga	gcaggggggtg	tgggcgcggc	ggtcgggctg	taaccccccc	3000
ctgcaccccc	ctccccgagt	tgctgagcac	ggcccggttc	cgggtgcggg	gctccgtgcg	3060
gggcgtggcg	cggggctcgc	cgtgcggggc	gggggggtgg	ggcaggtggg	gggtgcgggc	3120
ggggcggggg	cgcctcgggc	cggggagggc	tcgggggagg	ggcgcgggcg	ccccggagcg	3180
ccggcggtcg	tcgaggcgcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcgaga	3240
gggcgcaggg	acttcctttg	tcccaaactc	ggcgagcccg	aaatctggga	ggcgccggcg	3300
cacccctctc	agcgggcgcg	ggcgaagcgg	tgccggcgcc	gcaggaagga	aatgggcggg	3360
gagggccttc	gtgcgtcgcc	gcgcgcgcgt	ccccctctcc	atctccagcc	tcggggctgc	3420
cgcaggggga	cggctgcctt	cgggggggac	ggggcagggc	ggggttcggc	ttctggcggtg	3480
tgaccggcgg	ctctagatgc	atgctcgagc	ggccgcaggt	gtgatggata	tctgcagaat	3540
tcgccccttc	tagaatgggg	gtgcacgggtg	agtactcgcg	ggctggggcg	tcggcccgcg	3600
ccgggtccct	gtttgagcgg	ggatttagcg	ccccggctat	tgggccaggag	gtggctgggt	3660
tcaaggaccg	gcgacttgct	aaggaccctg	gaagggggag	gggggtgggg	tgccctccacg	3720
tgccagcggg	gacttggggg	agtccttggg	gatggcaaaa	acctgacctg	tgaaggggac	3780
acagtttggg	gggttgagggg	aagaagggtt	gggggttctg	ctgtgccagt	ggagagggaag	3840
ctgataagct	gataacctgg	gcgctggagc	caccacttat	ctgccagagg	ggaagcctct	3900
gtcacaccag	gattgaagtt	tgcccgagga	agtggatgct	ggtagctggg	gggtgggggtg	3960
gcacacggca	gcaggattga	atgaaggcca	gggaggcagc	acctgagtg	ttgcatgggt	4020
ggggacagga	aggacgagct	ggggcagaga	cgtggggatg	aagggaagctg	tccttcacaa	4080
gccacccttc	tcctccccc	cctgactctc	agcctggcta	tctgttctag	aatgtcctgc	4140
ctgggtgtgg	cttctcctgt	ccctgctgtc	gctccctctg	ggcctccag	tcctggggcg	4200
cccaccagc	ctcatctgtg	acagccgagt	cctggagagg	tacctcttgg	aggccaagga	4260
ggccgagaat	atcacgggtg	gaccccttcc	ccagcactt	ccacagaact	cacgctcagg	4320
gcttcaggga	actcctccca	gatccaggaa	cctggcactt	ggtttggggg	ggagttggga	4380
agctagacac	tgcccccccta	cataagaata	agtcctgggtg	ccccaaacca	tacctggaaa	4440
ctaggcaagg	agcaaaagcca	gcagatcccta	cgccctgtgg	gccagggcca	gagccttcag	4500
ggacccttga	ctccccgggc	tgtgtgcatt	tcagacgggc	tgtgctgaac	actgcagctt	4560

gaatgagaat	atcactgtcc	cagacaccaa	agttaatttc	tatgcctgga	agaggatgga	4620
ggtgagttcc	tttttttttt	tttttccctt	cttttggaga	atctcatttg	cgagcctgat	4680
tttggatgaa	agggagaatg	atcgagggaa	aggtaaaatg	gagcagcaga	gatgaggctg	4740
cctgggcgca	gaggctcacg	tctataatcc	caggctgaga	tggccgagat	gggagaattg	4800
cttgagccct	ggagtttcag	accaacctag	gcagcatagt	gagatcccc	atctctacaa	4860
acatttaaaa	aaatttagtca	ggtgaagtgg	tgcattggtg	tagtcccaga	tatttgggaag	4920
gctgaggcgg	gaggatcgct	tgagcccagg	aatttgaggc	tgcatgagc	tgtgatcaca	4980
ccactgcact	ccagcctcag	tgacagagtg	aggccctgtc	tcaaaaaaga	aaagaaaaaa	5040
gaaaaataat	gagggctgta	tggaaatacat	tcattattca	ttcactcact	cactcactca	5100
ttcattcatt	cattcattca	acaagtctta	ttgcatacct	tctgtttgct	cagcttggtg	5160
cttggggctg	cttgggggca	ggagggagag	ggtgacatgg	gtcagctgac	tcccagagtc	5220
cactccctgt	aggtcgggca	gcaggccgta	gaagtctggc	agggcctggc	cctgctgtcg	5280
gaagctgtcc	tgcgggggca	ggccctgttg	gtcaactctt	cccagccgtg	ggagcccctg	5340
cagctgcatg	tggataaagc	cgctcagtg	cttcgcagcc	tcaccactct	gcttcgggct	5400
ctgggagccc	aggtgagtag	gagcggacac	ttctgcttgc	cctttctgta	agaaggggag	5460
aagggctctg	ctaaggagta	caggaactgt	ccgtattcct	tccctttctg	tggcactgca	5520
gcgacctcct	gtttttctct	tggcagaagg	aagccatctc	ccctccagat	gcggcctcag	5580
ctgctccact	ccgaacaatc	actgctgaca	ctttccgcaa	actcttcoga	gtctactcca	5640
atttctcccg	gggaaagctg	aagctgtaca	caggggaggg	ctgcaggaca	ggggacagat	5700
gacgtacaag	taagaattca	ctcctcaggt	gcaggctgcc	tatcagaagg	tgggtggctg	5760
tgtggccaat	gccctggctc	acaaatacca	ctgagatctt	tttccctctg	ccaaaaatta	5820
tggggacatc	atgaagcccc	ttgagcatct	gacttctggc	taataaagga	aatttatattt	5880
cattgcaata	gtgtgttggg	attttttgtg	tctctcactc	ggaaggacat	atggcgggct	5940
aaatcattta	aaacatcaga	atgagtattt	ggttttagagt	ttggcaacat	atgccatatg	6000
ctggctgcca	tgaacaaagg	tggctataaa	gaggtcatca	gtatatgaaa	cagccccctg	6060
ctgtccattc	cttattccat	agaaaagcct	tgacttgagg	ttagattttt	tttatatttt	6120
gttttgtgtt	atttttttct	ttaacatccc	taaaattttc	cttacatgtt	ttactagcca	6180
gatttttctc	cctctcctga	ctactcccag	tcatagctgt	ccctcttctc	ttatgaagat	6240
ccctcgacct	gcagcccaag	cttgcatgcc	tgagggtcga	ctctagtggg	tcccccgccc	6300
cgtatccccc	agggtgtctgc	aggctcaaag	agcagcgaga	agcgttcaga	ggaaagcgat	6360
cccgtgccac	cttccccgtg	cccgggctgt	ccccgcacgc	tgcgggctcg	gggatgcccgg	6420
gggagcgccg	gagcccgagc	gagcccccgg	cggtctcgct	ctgcccccta	gcgggggggc	6480
gacgtaatta	catccctggg	ggctttgggg	gggggctgtc	cccgtgagcg	gatccgcggc	6540
cccgtatccc	ccaggtgtct	gcaggctcaa	agagcagcga	gaagcgttca	gaggaaagcg	6600
atcccgtgcc	acgtttcccc	tgccccggct	gtccccgcac	gctgcccggc	gcggggtcgc	6660
gggggagcgc	cggaccggag	cggagccccc	ggcggctcgc	tgctgcccc	tagcggggga	6720
gggacgtaat	tacatccctg	ggggctttgg	gggggggctg	tccccgtgag	cggatcccgc	6780
gccccgtatc	cccaggtgtc	ctgcaggctc	aaagagcagc	gagaagcgtt	cagaggaag	6840
cgatcccgtg	ccaccttccc	cgtgccccgg	ctgtccccgc	acgctgccgg	ctcggggatg	6900
cggggggagc	gcccagaccg	agcggagccc	cgggcggctc	gctgctgccc	cctagcgggg	6960
gagggagcta	attacatccc	tgggggcttt	gggggggggc	tgtccccgtg	agcggatccg	7020
cggccccgta	tccccaggtc	gtctgcaggg	tcaaagagca	gcgagaagcg	ttcagaggaa	7080
agcgatcccc	tgccaccttc	cccgtgcccc	ggctgtcccc	gcacgctgcc	ggctcgggga	7140
tgcgggggga	gcgcgggacc	ggagcggagc	cccgggcggc	tcgctgctgc	cccctagcgg	7200
gggagggacg	taattacatc	cctgggggct	ttgggggggg	gctgtccccg	tgagcggatc	7260
cgcgggcccc	tatccccag	gtgtctgcag	gctcaaagag	cagcgagaag	cgttcagagg	7320
aaagcgatcc	cgtgccacct	tccccgtgcc	cgggctgtcc	ccgcacgctg	ccggctcggg	7380
gatgcggggg	gagcgccgga	ccggagcggg	gccccgggcg	gctcgctgct	gccccctagc	7440
gggggagggg	cgtaattaca	tccctggggg	ctttgggggg	gggctgtccc	cgtgagcggg	7500
tccgcggccc	cgtatcccc	agggtgtctg	aggctcaaag	agcagcgaga	agcgttcaga	7560
ggaaagcgat	cccgtgccac	cttccccgtg	cccgggctgt	ccccgcacgc	tgcgggctcg	7620
gggatgcggg	gggagcgccg	gaccggagcg	gagccccggg	cggctcgctg	ctgcccccta	7680
gcgggggag	gacgtaatta	catccctggg	ggctttgggg	gggggctgtc	cccgtgagcg	7740
gatccgcggg	gctgcaggaa	ttcgtaatca	tggtcatagc	tgtttcctgt	gtgaaattgt	7800
tatccgctca	caattccaca	caacatacga	gccggaagca	taaagtgtaa	agcctggggg	7860
gcctaattgag	tgagctaact	cacattaatt	gcgttgctgt	cactgcccgc	tttccagctg	7920
ggaaacctgt	cgtgccagct	gcattaatga	atcgcccaac	gcgcggggag	aggcgggttg	7980
cgtattgggc	gctcttccgc	ttcctcgctc	actgactcgc	tgcgctcggt	cgttcgggct	8040
cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	atcaggggat	8100
aacgcaggaa	agaataatgt	agcaaaaggc	ccaggaaccg	ccaggaaccg	taaaaaggcc	8160
gcgttgctgg	cgtttttcca	taggctccgc	ccccctgacg	agcatcacia	aaatcgacgc	8220
tcaagtacga	ggtggcgaaa	cccgcacagg	ctataaagat	accaggcggt	tccccctgga	8280
agctccctcg	tgcgctctcc	tgctccgacc	ctgacgctta	ccggataacct	gtccgctttt	8340
ctcccttcgg	gaagcgtggc	gctttctcat	agctcacgct	gtaggtatct	cagttcgggtg	8400
taggtcgttc	gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	8460
gccttatccg	gttaactatc	tcttgagtc	accccggtaa	gacacgactt	atcgccactg	8520
gcagcagcca	ctggtaacag	gattagcaga	gcgaggtatg	taggcgggtgc	tacagagttc	8580

ttgaagtgggt	ggcctaacta	cggtctacact	agaaggacag	tatttggtat	ctgcgctctg	8640
ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	8700
gctggtagcg	gtgggtttttt	tgtttgcaag	cagcagatta	cgcgagaaa	aaaaggatct	8760
caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	agtggaacga	aaactcacgt	8820
taagggatttt	tggtcatgag	attatcaaaa	aggatcttca	cctagatcct	tttaaatata	8880
aatgaagtt	ttaaatcaat	ctaaagtata	tatgagtaaa	cttgggtctga	cagttaccaaa	8940
tgcttaataca	gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcate	catagttgcc	9000
tgactccccg	tcgtgtagat	aactacgata	cgggagggct	taccatctgg	ccccagtgct	9060
gcaatgatac	cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	9120
gccggaaggg	ccgagcgag	aagtggctct	gcaactttat	ccgcctccat	ccagttctatt	9180
aattgttgcc	gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	caacgttggt	9240
gccattgcta	caggcatcgt	ggtgtcacgc	tcgtcgtttg	gtatggcttc	attcagctcc	9300
ggttcccaac	gatcaaggcg	agttacatga	tccccatgt	tgtagcaaaa	agcgggttagc	9360
tccttcggtc	ctccgatcgt	tgtagaaggt	aagttggccg	cagtggttatc	actcatgggt	9420
atggcagcac	tgcatatttc	tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	9480
ggtgagtact	caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	9540
ccggcgtaaa	tacgggataa	taccgcgcca	catagcagaa	ctttaaagt	gctcatcatt	9600
ggaaaaacgtt	cttcggggcg	aaaactctca	aggatcttac	cgctgttgag	atccagttcg	9660
atgtaaccca	ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	cagcgtttct	9720
gggtgagcaa	aaacaggaag	gcaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	9780
tggtgaatac	tcatactctt	cttttttcaa	tattattgaa	gcatttatca	gggttattgt	9840
ctcatgagcg	gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	ggttccgcgc	9900
acatttcccc	gaaaagtgcc	acctgacgta	gttaacaaaa	aaaagcccg	cgaagcgggc	9960
tttattacca	agcgaagcgc	cattcgccat	tcaggctgcg	caactgttgg	gaagggcgat	10020
cggtgcgggc	ctcttcgcta	ttacgccagc	tggcgaaagg	gggatgtgct	gcaaggcgat	10080
taagttgggt	aacgccaggg	ttttcccagt	cacgacgttg	taaaacgacg	gccagtccgt	10140
aatacgactc	acttaaggcc	ttgactagag	ggtcgacggg	atacagacat	gataagatac	10200
attgatgagt	ttggacaaac	cacaactaga	atgcagtgaa	aaaaatgctt	tatttgtgaa	10260
atttgtgatg	ctattgcttt	atttgttaacc	attataagct	gcaataaaca	agttgggggtg	10320
ggcgaagaac	tcacgcatga	gatccccgcg	ctggaggatc	atccagccgg	cgccccggaa	10380
aacgattccg	aagcccaacc	tttcatagaa	ggcggcggtg	gaatcgaaat	ctcgtagcac	10440
gtgtcagtc	tgctcctcgg	ccacgaagtg	cacg			10474

<210> 126

<211> 6119

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeoeGFP Plasmid

<400> 126

cagttgcccgg	ccgggtcgcg	cagggcgaa	tcccccccc	acggctgctc	gccgatctcg	60
gtcatggccg	gcccggaggc	gtcccgaag	ttcgtggaca	cgacctccga	ccactcggcg	120
tacagctcgt	ccaggccgcg	caccacacac	caggccaggg	tggtgtccgg	caccacctgg	180
tcctggaccg	cgctgatgaa	cagggctcacg	tcgtcccggg	ccacaccggc	gaagtgcgtc	240
tcacgaagt	ccggggagaa	cccagcccg	tcgggtccaga	actcgaccgc	tccggcgacg	300
tcgcgcgcgg	tgagcaccgg	aacggcactg	gtcaacttgg	ccatggatcc	agatttcgct	360
caagttagta	taaaaaagca	ggcttcaatc	ctgcagagaa	gcttgggctg	caggtcgagg	420
gatcttcata	agagaagagg	gacagctatg	actgggagta	gtcaggagag	gaggaaaaat	480
ctggctagta	aaacatgtaa	ggaaaatttt	agggatgta	aagaaaaaaa	taacacaaaa	540
caaaatataa	aaaaaatcta	acctcaagtc	aaggcttttc	tatggaataa	ggaatggaca	600
gcaggggggt	gtttcatata	ctgatgacct	ctttatagcc	acctttgttc	atggcagcca	660
gcataatggca	tatgttgcca	aactctaaac	caaataactca	ttctgatgtt	ttaaatgatt	720
tgccctccca	tatgtccttc	cgagtgcagc	acacaaaaaa	ttccaacaca	ctattgcaat	780
gaaaaataat	ttcctttatt	agccagaagt	cagatgctca	aggggcttca	tgatgtcccc	840
ataatttttg	gcagagggga	aaagatctca	gtggtatttg	tgagccaggg	catgggccac	900
accagccacc	accttctgat	aggcagcctg	cacctgagga	gtgaattctt	acttgtacag	960
ctcgtccatg	ccgagagtga	tcccggcgcc	ggtcacgaac	tccagcagga	ccatgtgac	1020
gcgcctctcg	ttgggggtct	tgctcagggc	ggactgggtg	ctcaggtagt	ggttgtcggg	1080
cagcagcacg	gggcccgtgc	cgatgggggt	gttctgctgg	tagtgggtcg	cgagctgcac	1140
gctgccgtcc	tcgatgttgt	ggcggatctt	gaagttcacc	ttgatgccgt	tcttctgctt	1200
gtcgggcagc	atatagactg	tggtgctgtt	gtagttgtac	tccagcttgt	gccccaggat	1260
ggtgccgtcc	tccttgaaagt	cgatgcccct	cagctcgatg	cggttcacca	gggtgtcgcc	1320
ctcgaacttc	acctcggcgc	gggtcttgtta	gttgccgtcg	tccttgaaaga	agatgggtgc	1380
ctcctggacg	tagccttcgg	gcattggcga	ctggaagaag	tcgtgctgct	tcattgtggtc	1440
gggggtacgg	ctgaagcact	gcacgcccga	ggtcaggggtg	gtcacgaggg	tgggcccagg	1500

cacgggcagc	ttgcgggtgg	tgcagatgaa	cttcagggtc	agcttgccgt	aggtggcatc	1560
gccctcgccc	tgcggggaca	cgctgaactt	gtggccgttt	acgtcgccgt	ccagctcgac	1620
caggatgggc	accaccccg	tgaacagctc	ctcgcccttg	ctcaccatgg	tggcgaaattc	1680
tttgccaaaa	tgatgagaca	gcacaacaac	cagcacgttg	cccaggagct	gtaggaaaaa	1740
gaagaaggca	tgaacatggt	tagcagaggc	tctagagccg	ccggtcacac	gccagaagcc	1800
gaaccccgcc	ctgccccgtc	ccccccgaag	gcagccgtcc	ccctgcggca	gccccgaggc	1860
tggagatgga	gaaggggacg	gcggcgccgc	gacgcacgaa	ggccctcccc	gccccatttc	1920
ttcctgccc	cgccgcaccc	cttcgcccgc	cccgctaga	gggggtgccc	cgcgccctcc	1980
cagatttcgg	ctccgccaga	tttgggacaa	aggaagtccc	tgcgccctct	cgcacgatta	2040
ccataaaagg	caattggctgc	ggctcgccgc	gcctcgacag	ccgccggcgc	tccggggccc	2100
ccgcgccct	cccccgacc	ctccccggcc	cgaggcgccc	ccgccccgcc	cggcaccccc	2160
acctgccgcc	accccccgcc	cggcacggcg	agccccgcgc	cacgccccgc	acggagcccc	2220
gcacccgaag	ccggggccgtg	ctcagcaact	cggggagggg	ggtgcagggg	ggggttacag	2280
cccgaaccgc	gcgcccacac	ccccccacgc	ccccccacgc	cacacacccc	gcacgcagcc	2340
tttgttcccc	tgcagcccc	cccgccacgc	ggggcaccgc	ccccggccgc	gctccccctc	2400
cgccacgcgc	gagcgcacaa	agccccgcgc	cgcgcccgca	gcgctcacag	ccgccgggca	2460
gcgcggggcg	gcgcggcgcc	tccccacgc	cacacacacg	cacgcacccc	ccgagccgct	2520
cccccccgca	caaagggccc	tccccggagcc	ctttaaggct	ttcacgcagc	cacagaaaag	2580
aaacagagcc	tcattaaacc	aagcgctaatt	tacagcccg	aggagaagg	ccgtcccgcc	2640
cgctcacctg	tgggagagac	gcggtcagtc	agagccgggg	cgggcgggcg	gaggcgggcg	2700
ggagcggggc	acggggcgaa	ggcaacgcag	cgactccgc	ccgccgcgcg	cttcgctttt	2760
tatagggcgc	ccgcgcgcgc	cgccctcgcca	taaaaggaaa	ctttcgaggc	gcgcccgtct	2820
gattggctgc	cgccgcacct	ctccgccctc	cccccccg	ccctcgccc	cgccccgcc	2880
cgccggcgcc	gcgccccccc	cccccccgcc	cccatcgctg	cacaaaaata	taaaaaata	2940
aataaataca	aaattggggg	tggggagggg	ggggagatgg	ggagagtga	gcagaacgtg	3000
gggctcacct	gcacccatgg	taatagcgat	gactaatacg	tagatgtact	gccaagttag	3060
aaagtcccat	aaggtcatgt	actgggcata	atgccaggcg	ggccatttac	cgctcattgac	3120
gtcaatagg	ggcgtaactg	gcatatgata	cacttgatgt	actgccaaat	gggcagttta	3180
ccgtaaatag	tccacccatt	gacgtcaatg	gaaagtcct	attggcgcta	ctatgggaac	3240
atacgtcatt	attgacgtca	atgggcgggg	gtcggtgggc	ggtcagccag	gcggggccatt	3300
taccgtaagt	tatgtaacgc	gggaactccat	atatgggcta	tgaactaatg	accccgtaatt	3360
atagctgttt	cctgtgtgaa	attgttatcc	gtccacaatt	tcgaattcgt	aatcatggtc	3420
aagcataaag	tgtaaagcct	gggggtgccta	atgagttagc	ccacacacaa	tacgagccgg	3480
gcgctcactg	cccgctttcc	agtcgggaaa	cctgtcgtgc	taactcacat	taattgcgtt	3540
ccaacgcgcg	gggagaggcg	gtttgcgtat	tgggcgtctc	cagctgcatt	aatgaatcgg	3600
ctcgctgcgc	tgggtcgctc	ggctgcggcg	agcggtatca	tccgcttcc	cgctcactga	3660
acggttatcc	acagaatcag	gggataacgc	aggaaagaa	gctcactcaa	aggcggtaat	3720
aaaggccagg	aaccgtaaaa	aggccgcgtt	gctggcgctt	atgtgagcaa	aaggcgagca	3780
tgacgagcat	cacaaaaaatc	gacgctcaag	tcagaggtgg	ttccataggc	tccgcccccc	3840
aagataccag	gcgtttcccc	ctggaagctc	cctcgtgcgc	cgaaaaccga	caggactata	3900
gcttacccga	tacctgtccg	cccttctccc	ttcgggaagc	tctcctgttc	cgaccctgcc	3960
acgctgtagg	tatctcagtt	cggtgtagg	cgttcgctcc	gtggcgcttt	ctcatagctc	4020
acccccggtt	cagccccgacc	gctgcgcctt	atccggtaac	aagctgggct	gtgtgcacga	4080
ggtaagacac	gacttatcgc	cactggcagc	agccactgg	tatcgtcttg	agtcacaccc	4140
gtatgtaggc	gggtgtacag	agttcttgaa	gtgggtggct	aacaggatta	gcagagcgag	4200
gacagtat	gggtatctgc	ctctgctgaa	gccagttacc	aactacggct	acactagaag	4260
ctcttgatcc	ggcaaaacaaa	ccaccgctgg	tagcgggtgg	ttcggaaaaa	gagttggtag	4320
gattacgcgc	agaaaaaaag	gatctcaaga	agatcccttg	ttttttgttt	gcaagcagca	4380
cgctcagtg	aacgaaaaact	cacgttaagg	gattttgggtc	atcttttcta	cggggtctga	4440
cttcacctag	atccttttaa	attaaaaatg	aagttttaaa	atgagattat	caaaaaaggat	4500
gtaaacttgg	tctgacagtt	accaatgctt	aatcagtgag	tcaatctaaa	gtatatatga	4560
tctatttcgt	tcattccatag	ttgctgact	ccccgctcgt	gcacctatct	cagcgatctg	4620
gggcttacca	tctggcccca	gtgctgcaat	gataccgcga	tagataacta	cgatacggga	4680
agatttatca	gcaataaacc	agccagccgg	aaggcccgag	gacccacgct	caccggctcc	4740
tttatccgcc	tccatccagt	ctattaattg	ttgcccggaa	cgagaaagtg	gtcctgcaac	4800
agttaatagt	ttgcgcaacg	ttgttgccat	tgctacaggc	gctagagtaa	gtagttcgcc	4860
gtttgggtatg	gcttcattca	gctccggttc	ccaacgatca	atcgtgggtg	cacgctcgtc	4920
catgttgtgc	aaaaaagcgg	ttagctcctt	cggtcctccg	aggcgagtta	catgatcccc	4980
ggccgagctg	ttatcactca	tggttatggc	agcactgcac	atcggtgtca	gaagtaagtt	5040
atccgtaaga	tgcttttctg	tgactgggtga	gtactcaacc	aattctctta	ctgtcagccc	5100
tatgcggcga	ccgagttgct	cttgcccggc	gtcaataccg	aagtcattct	gagaatagtg	5160
cagaacttta	aaagtgtcca	tttgagatcca	acgtttctcg	gataataccg	cgccacatag	5220
cttaccgctg	ttgagatcca	gttcgatgta	acccactcgt	gggcgaaaaa	tctcaaggat	5280
atcttttact	ttccaccagcg	tttctgggtg	agcaaaaaca	gcacccaact	gatcttcagc	5340
aaaggggaata	aggcgacac	ggaaatgttg	aataactata	ggaaggcaaa	atgccgcaaa	5400
ttgaagcatt	tatcaggggt	attgtctcat	gagcggatac	ctcttccttt	ttcaatatga	5460
				atatttgaat	gtatttagaa	5520

aaataaacaa	ataggggttc	cgcgcacatt	tccccgaaaa	gtgccacctg	acgtagttaa	5580
caaaaaaaaaag	cccgccgaag	cgggctttat	taccaagcga	agcgccattc	gccattcagg	5640
ctgcgcaact	gttggaagg	gcgatcgggtg	cgggcctctt	cgctattacg	ccagctggcg	5700
aaagggggat	gtgctgcaag	gcgattaagt	tgggtaacgc	cagggttttc	ccagtcacga	5760
cgttgtaaaa	cgacggccag	tccgtaatac	gactcactta	aggccttgac	tagagggctg	5820
acggtataca	gacatgataa	gatacattga	tgagtttgga	caaaccacaa	ctagaatgca	5880
gtgaaaaaaa	tgctttattt	gtgaaatttg	tgatgctatt	gctttatttg	taaccattat	5940
aagctgcaat	aaacaagttg	gggtgggcga	agaactccag	catgagatcc	ccgcgctgga	6000
ggatcatcca	gccggcgctc	cggaaaacga	ttccgaagcc	caacctttca	tagaaggcgg	6060
cgggtggaatc	gaaatctcgt	agcacgtgtc	agtctgtctc	ctcggccacg	aagtgcacg	6119

<210> 127

<211> 5855

<212> DNA

<213> Artificial Sequence

<220>

<223> pCXLamInt Plasmid (Wildtype Integrase)

<400> 127

gtcgacattg	attattgact	agttattaat	agtaatcaat	tacgggggtca	ttagttcata	60
gcccataat	ggagttccgc	gttacataac	ttacggtaaa	tggcccgccct	ggctgaccgc	120
ccaacgaccc	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	acgccaatag	180
ggactttcca	ttgacgtcaa	tgggtggact	atttacggta	aactgccac	ttggcagtac	240
atcaagtgt	tcatatgcca	agtacgcccc	ctattgacgt	caatgacggt	aaatggcccg	300
cctggcatta	tgcccagtac	atgaccttat	gggactttcc	tacttggcag	tacatctacg	360
tattagtc	cgctattacc	atgggtcgag	gtgagcccca	cgttctgctt	cactctcccc	420
atctcccccc	cctccccacc	cccaattttg	tattttatta	ttttttaatt	atthttgtgca	480
gcgatggggg	cggggggggg	ggggggcgcg	gccaggcggg	gcggggcggg	gcgagggggc	540
gggcggggcg	aggcgagag	gtgcggcggc	agccaatcag	agcgggcgcg	tccgaaagt	600
tccttttatg	gcgagggcgc	ggcgggcggc	gccctataaa	aagcgaagcg	cgcgggcggc	660
gggagtcgct	gcgttgccct	cgccccgtgc	cccgtccgc	gccgcctcgc	gccgcccgcc	720
ccggctctga	ctgaccgcgt	tactcccaca	ggtgagcggg	cgggacggcc	cttctcctcc	780
gggctgta	tagcgcttgg	tttaattgacg	gctcgcttct	tttctgtggc	tgctgaaag	840
ccttaaagg	ctccgggagg	gccctttgtg	cgggggggag	cggtcggggg	ggtgcgtgcg	900
tgtgtgtgtg	cgtggggagc	gccgcgtgcg	gcccgcgctg	cccgccggct	gtgagcgctg	960
cgggcgcggc	gcggggcctt	gtgcgctccg	cgtgtgcgcg	aggggagcgc	ggccgggggc	1020
ggtgccccgc	ggtgcggggg	ggctgcgagg	ggaacaaagg	ctgcgtgcgg	ggtgtgtgcg	1080
tgggggggtg	agcaggggtg	gtgggcgcgg	cggtcgggct	gtaaccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cgccccggct	tcgggtgcgg	ggctccgtgc	ggggcggtgc	1200
gcgggggctc	ccgtgcgcgg	cgggggggtg	cggcaggtgg	gggtgcgcgg	cggggcgggg	1260
ccgcctcggg	ccggggaggg	ctcggggagg	gggcgcggcg	gcccgggagc	gcggggcgct	1320
gtcgaggcgc	ggcgagccgc	agccattgccc	ttttatggta	atcggtgcgag	agggcgcgag	1380
gacttccttt	gtcccaaattc	tggcgaggcc	gaaatctggg	aggcgcccgcc	gcacccccct	1440
tagcggtgcg	gggcgaagcg	gtgcggcgcc	ggcaggaagg	aaatgggcgg	ggagggcgct	1500
cgtgcgtcgc	cgcgcgcgcg	tccccttctc	catctccagc	ctcggggctg	ccgcaggggg	1560
acggctgcct	tcggggggga	cggggcgagg	cgggggttcg	cttctggcgt	gtgaccggcg	1620
gctctagagc	ctctgctaac	catgttcattg	ccttctctct	tttctacag	ctcctgggca	1680
acgtgctggt	tgttgtgtcg	tctcatcatt	ttggcaaa	attcatggga	agaaggcgaa	1740
gtcatgagcg	ccgggattta	ccccctaacc	tttatataag	aaacaatgga	tattactgct	1800
acagggaccc	aaggacgggt	aaagagtttg	gattaggcag	agacaggcga	atcgcaatca	1860
ctgaagctat	acaggccaac	attgagttat	tttcaggaca	caaacacaa	cctctgacag	1920
cgagaatcaa	cagtgataat	tccgttacgt	tacattcatg	gcttgatcgc	tacgaaaaaa	1980
tcctggccag	cagaggaatc	aagcagaaga	cactcataaa	ttacatgagc	aaaattaaag	2040
caataaggag	gggtctgcct	gatgctccac	ttgaagacat	caccacaaaa	gaaattgcgg	2100
caatgctcaa	tggatacata	gacgagggca	aggcggcgtc	agccaagtta	atcagatcaa	2160
cactgagcga	tgcatccgga	gaggcaatag	ctgaaggcca	tataacaaca	aaccatgtcg	2220
ctgccactcg	cgcagcaaaa	tcagaggtaa	ggagatcaag	acttacggct	gacgaatacc	2280
tgaaaaattta	tcaagcagca	gaatcatcac	catgttggct	cagacttgca	atggaactgg	2340
ctgttggttac	cgggcaacga	gttggtgatt	tattcgaaat	gaagtggctc	gatatcgtag	2400
atggatatct	ttatgtcgag	caaagcaaaa	caggcgtaaa	aattgccatc	ccaacagcat	2460
tgcatattga	tgtctcgcga	atatcaatga	aggaacaact	tgataaatgc	aaagagattc	2520
ttggcggaga	aaccataatt	gcattctactc	gtcgcgaacc	gcttccatcc	ggcagagtat	2580
caagggtattt	tatgcgcgca	cgaaaagcat	caggctcttc	cttcgaaggg	gatccgccta	2640
cctttcacga	tgttcgcag	ttgtctgcaa	gactctatga	gaagcagata	agcgataagt	2700
ttgtcaaca	ctttctcggg	cataagtcgg	accattggc	atcacagtat	cgtgatgaca	2760
gaggcagggg	gtgggacaaa	attgaaatca	aataagaatt	cactcctcag	gtgcaggctg	2820

cctatcagaa	ggtggtggct	ggtgtggcca	atgccctggc	tcacaaatac	cactgagatc	2880
tttttccctc	tgccaaaaat	tatggggaca	tcatgaagcc	ccttgagcat	ctgacttctg	2940
gctaataaag	gaaattttatt	ttcattgcaa	tagtgtgttg	gaattttttg	tgtctctcac	3000
tcggaaggac	atatggggagg	gcaaatcatt	taaaacatca	gaatgagtat	ttggtttaga	3060
gtttggcaac	atatgccata	tgctggctgc	catgaacaaa	ggtggctata	aagagggtcat	3120
cagtatatga	aacagccccc	tgctgtccat	tccttattcc	atagaaaagc	cttgacttga	3180
ggttagattt	ttttttatatt	ttgtttttgtg	ttattttttt	ctttaacatc	cctaaaattt	3240
tccttacatg	ttttactagc	cagatttttct	ctcctctcct	gactactccc	agtcatagct	3300
gtccctcttc	tcttatgaag	atccctcgac	ctgcagccca	agcttggcgt	aatcatggtc	3360
atagctgttt	cctgtgtgaa	attgttatcc	gctcacaatt	ccacacaaca	tacgagccgg	3420
aagcataaag	tgtaaagcct	gggggtgccta	atgagtgagc	taactcacat	taattgcgtt	3480
gcgctcactg	cccgttttcc	agtcgggaaa	cctgtcgtgc	cagcggatcc	gcattctcaat	3540
tagtcagcaa	ccatagtccc	gcccttaact	cgcccataac	cgcccataac	tccgccagct	3600
tccgcccat	ctccgcccc	tggtgacta	atttttttta	tttatgcaga	ggccgaggcc	3660
gcctcggcct	ctgagctatt	ccagaagtag	tgaggaggct	tttttggagg	cctaggcttt	3720
tgcaaaaagc	taacttggtt	attgcagctt	ataatggtta	caaataaagc	aatagcatca	3780
caaatttcac	aaataaagca	tttttttcac	tgcatctag	ttgtggtttg	tccaaactca	3840
tcaatgtatc	ttatcatgtc	tggtatccgc	gcattaatga	atcggccaac	gcgcggggag	3900
aggcggtttg	cgtattgggc	gctctccgc	actgactcgc	actgactcgc	tgccctcggt	3960
cgttcggctg	cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	4020
atcaggggat	aacgcaggaa	agaacatgtg	agcaaaaagg	cagcaaaaagg	ccaggaaccc	4080
taaaaaggcc	gcgttgctgg	cgtttttcca	taggctccgc	ccccctgacg	agcatcacaa	4140
aaatcgacgc	tcaagtccga	gggtggcgaaa	cccgacagga	ctataaagat	accaggcggt	4200
tccccctgga	agctccctcg	tgccctctcc	tggtccgacc	ctgccgctta	ccggatacct	4260
gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctcaa	tgctcacgct	gtaggatatc	4320
cagttcgggtg	taggtcgttc	gctccaagct	ggcgtgtgtg	cacgaacccc	ccgttcagcc	4380
cgaccgctgc	gccttatccg	gtaactatcg	tcttgagctc	aaccgggtta	gacacgactt	4440
atcgccactg	gcagcagcca	ctggtaacag	gattagcaga	gcgaggatat	taggcgggtg	4500
tacagagttc	ttgaagtggg	ggcctaacta	cggctacact	agaaggacag	tatttggtat	4560
ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	4620
acaaaccacc	gctggtagcg	gtggtttttt	tggttgcaag	cagcagatta	cgcgagaaa	4680
aaaaggatct	caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	agtggaaacga	4740
aaactcacgt	taagggattt	tggtcatgag	attatcaaaa	aggatcttca	cctagatcct	4800
tttaaattaa	aatgaagtt	ttaaatcaat	ctaaagtata	tatgagtaaa	cttgggtctga	4860
cagttaccaa	tgcttaatca	gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcatc	4920
catagttgcc	tgactccccg	tcgtgtagat	aactacgata	cgggagggct	taccatctgg	4980
cccagtgct	gcaatgatac	cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	5040
aaaccagcca	gccggaagg	ccgagcgcag	aagtggctct	gcaactttat	ccgcctccat	5100
ccagtctatt	aattgttgcc	gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	5160
caacgttggt	gccattgcta	caggcatcgt	gggtgcacgc	tcgtcggttg	gtatggcttc	5220
attcagctcc	ggttcccaac	gatcaaggcg	agttacatga	tccccatgt	tgtgcaaaaa	5280
agcggtagc	tccttcggtc	ctccgatcgt	tgtcagaagt	aagttggccg	cagtggtatc	5340
actcatgggt	atggcagcac	tgcataattc	cttctactgtc	atgccatccg	taagatgctt	5400
ttctgtgact	gttgagtact	caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	5460
ttgctcttgc	ccggcgctcaa	tacgggataa	taccgcgcca	catagcagaa	ctttaaaagt	5520
gctcatcatt	ggaaaacgtt	cttcggggcg	aaaactctca	aggatcttac	cgctggtgag	5580
atccagttcg	atgtaaccca	ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	5640
cagcgtttct	gggtgagcaa	aaacaggaag	gcaaaatgcc	gcaaaaaagg	gaataagggc	5700
gacacggaaa	tggtgaatac	tcatactctt	cctttttcaa	tattattgaa	gcatttatca	5760
gggttattgt	ctcatgagcg	gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	5820
ggttccgcgc	acatttcccc	gaaaagtgcc	acctg			5855

<210> 128

<211> 303

<212> DNA

<213> Artificial Sequence

<220>

<223> Human FER-1 Promoter

<400> 128

tccatgacaa	agcacttttt	gagcccaagc	ccagcctagc	tcgagctaaa	cgggcacaga	60
gacgccaccg	ctgtcccaga	ggcagtcggc	taccgggtccc	cgctcccag	ctccgccaga	120
gcgcgcgagg	gcctccagcg	gccgcctctc	ccccacagca	ggggcgggg	cccgcgcccc	180
ccggaaggag	cgggctcggg	gcgggcggcg	ctgattggcc	ggggcggg	tgacgccgac	240
gcggctataa	gagaccacaa	gcgaccgcga	gggcagagc	ttcttcggcg	agagtcgggt	300
acc						303

<210> 129
<211> 6521
<212> DNA
<213> Artificial Sequence

<220>
<223> pIRES-BSR Plasmid

<400> 129
tcaatattgg ccattagcca tattattcat tgggttatata gcataaatca atattgggcta 60
ttggccattg catacgttgt atctatatca taatatgtac atttatattg gctcatgtcc 120
aatatgaccg ccatgtttggc attgattatt gactagttat taatagtaat caattacggg 180
gtcattagtt catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc 240
gcctggctga ccgcccacag acccccggccc attgacgtca ataatgacgt atgttcccat 300
agtaacgcca atagggactt tccattgacg tcaatgggtg gagtatttac ggtaaaactgc 360
ccacttggca gtacatcaag tgtatcatat gccaaagtccg cccctatttg acgtcaatga 420
cggtaaatgg ccgcctgggc attatgcccc gtacatgacc ttacgggact ttccctacttg 480
gcagtacatc tacgtattag tcatcgctat taccatgggt atgcgggtttt ggcagtacac 540
caatggggcg ggtatgcacg ttgactcacg gggatttcca agtctccacc ccattgacgt 600
caatgggagt ttgttttggc accaaaatca accgggacttt ccaaaatgtc gtaacaactg 660
cgatcgcccc ccccggttgac gcaaatgggc ggtaggcggtg tacgggtggga ggtctatata 720
agcagagctc gtttagtgaa ccgtcagatc actagaagct ttattgoggt agtttatcac 780
agttaaattg ctaacgcagt cagtgccttc gacacaacag tctcgaactt aagctgcagt 840
gactctctta aggtagcctt gcagaagttg gtctgtaggc actgggcagg taagtatcaa 900
ggttacaaga cagggtttaag gagaccaata gaaactgggc ttgtcgagac agagaagact 960
cttgcgtttc tgataggcac ctattggtct tactgacatc cactttgcct ttctctccac 1020
aggtgtccac tcccagttca attacagctc ttaaggctag agtacttaat acgactcact 1080
ataggctagc ctcgagaatt caccgctcga gcatgcactt agggcgggcca attccgcccc 1140
tctccctccc cccccctaa cgttactggc cgaagccgct tgggaataagg ccggtgtgcg 1200
tttgtctata tgtgattttc caccatattg ccgtcttttg gcaatgtgag ggcccgga 1260
cctggccctg tcttcttgac gagcattcct aggggtcttt cccctctcgc caaaggaatg 1320
caaggtctgt tgaatgtcgt gaaggaagca gtctctctgg aagcttcttg aagacaaa 1380
acgtctgtag cgaccctttg caggcagcgg aacccccac ctggcgacag gtgcctctgc 1440
ggccaaaagc cagctgtata agatacacct gcaaaggcgg cacaacccca gtgccacgtt 1500
gtgagttgga tagttgtgga aagagtcaaa ttgctctcct caagcgtatt caacaagg 1560
ctgaaggatg ccagaagggt acccatttgt atgggactct atctggggcc tgggtgcaca 1620
tgctttacat gtgttttagtc gaggttaaaa aaacgtctag gcccccgaa ccacggggg 1680
gtggttttcc ttgaaaaaac acgatgataa cgttgccaca accaccatg aaaacatt 1740
acatttctca acaagatcta gaattagtag aagtagcgac agagaagatt acaatgct 1800
atgaggataa taaacatcat gtgggagcgg caattcgtac gaaaacagga gaaatcatt 1860
cggcagtaac agtattcgaat ggacaaaagg attttgacac gattgtagct gttagacac 1920
ttggtagtgc cgaagttagt agaagtattc ttgtgttaat agaaatgaat ccgaaatt 2040
cttattctga cgaagttagt ccagatttgt tcttctctgg ttaatgcttc gagcagac 2100
agttgatttc gattgaagaa ctcatccac ttagtgaggg ttaatgcttc gagcagac 2160
ataccaagct tggcggggcg ccgcttccct tttagtgaggg ttaatgcttc gagcagac 2220
gataagatac attgatgagt ttggacaaac cacaactaga atgcagtga aaaaatgct 2280
tatttgtgaa atttgtgatg ctattgcttt atttghtaacc attataagct gcaataaaca 2340
agttaacaac aacaattgca ttcattttat gtttcagggt cagggggaga tgtgggaggt 2400
tttttaagc aagtaaaacc tctacaaatg ttggtaaaatc cgataaggat cgatccggg 2460
tggcgttaata gccaagaggc ccgcaccgat cgcccttccc aacagttgag cagcctga 2520
ggcgaatgga ccgcacctgt agcggcgcat taagcgcggc ggggtgtggt gttacgcg 2580
gcgtgaccgc tacacttgcc agcgccttag cgcccgctcc ttctcgtttc ttcccttcc 2640
ttctcgccac gttcgccggc ttccccgctc aagctctaaa tcgggggctc cctttagggt 2700
tccgatttag agctttacgg cacctcgacc gcaaaaaact tgatttgggt gatggttc 2760
gtagtgggccc atcgccctga tagacggttt ttccgctttt gacgttggag tccacgtt 2820
ttaatagtgg actcttgttc caaactggaa caaactcaa ccttatctcg gtctattct 2880
ttgatttata agggattttg ccgatttcgg cctattgggt aaaaaatgag ctgattta 2940
aaatatttaa cgcgaatttt aacaaaatat taacgtttac aatttcgct gatgcggt 3000
tttctcctta cgcactgtg cggtatttca caccgcatac gcggatctgc gcagcacc 3060
ggcctgaaat aacctctgaa agaggaactt ggttaggtac cttctgaggg ggaaagaa 3120
agctgtggaa tgtgtgtcag ttaggggtgt gaagtcgcc gaaccagggt tggaaagtcc 3180
gtatgcaaa ctagtctctc aattagtcag aattagtcag agcaaccata gtcggccccc 3240
cagcaggcag aagtagtcaa agcatgcac ccaactccgc ccatctccg cccattgg 3300
taactccgccc ctaactccgc gcagaggcgg aggcgcgctc ggcctctgag ctattccaga 3360
gactaatttt ttttatttat ggaggccctag gcttttgcaa aaagcttgat tcttctgaca 3420
agtagtgagg aggcctttttt

-104-

caacagtctc	gaacttaagg	ctagagccac	catgattgaa	caagatggat	tgcacgcagg	3540
ttctccggcc	gcttgggtgg	agaggctatt	cggctatgac	tgggcacaa	agacaatcgg	3600
ctgctctgat	gccgccgtgt	tccggctgtc	agcgcagggg	cgcccgggtc	tttttgtcaa	3660
gaccgacctg	tccggtgccc	tgaatgaact	gcaggacgag	gcagcgcggc	tatcgtggct	3720
ggccacgacg	ggcgtttcctt	gcgcagctgt	gctcgacgtt	gtcactgaag	cggaaggga	3780
ctggctgcta	ttggggcgaag	tgccggggca	ggatctcctg	tcactctacc	ttgctcctgc	3840
cgagaaaagta	tccatcatgg	ctgatgcaat	gcggcggtg	catacgcttg	atccggctac	3900
ctgcccattc	gaccaccaag	cgaaacatcg	catcgagcga	gcacgtactc	ggatggaagc	3960
cggctcttgtc	gatcaggatg	atctggacga	agagcatcag	gggctcgcgc	cagccgaact	4020
gttcgccagg	ctcaaggcgc	gcatgccga	cggcgaggat	ctcgtcgtga	cccatggcga	4080
tgctgcttg	ccgaatatca	tgggtgaaaa	tggccgcttt	tctggattca	tcgactgtgg	4140
ccggctgggt	gtggcggacc	gctatcagga	catagcgttg	gctacccgtg	atattgctga	4200
agagcttggc	ggcgaatggg	ctgaccgctt	cctcgtgctt	tacggtatcg	ccgctcccga	4260
ttcgcagcgc	atcgccctct	atcgccctct	tgacgagttc	ttctgagcgg	gactctgggg	4320
ttcgaaatga	ccgaccaagc	gacgcccac	ctgccatcac	gatggccgca	ataaaatc	4380
tttattttca	ttacatctgt	gtgttggttt	tttgtgtgaa	tcgatagcga	taaggatccg	4440
cgtatggtgc	actctcagta	caatctgctc	tgatgccgca	tagttaagcc	agccccgaca	4500
cccgccaaaca	cccgtcgacg	cgccctgacg	ggcttgtctg	ctcccggcat	ccgcttacag	4560
acaagctgtg	accgtctccg	ggagctgcat	gtgtcagagg	ttttcacctg	catcacgcaa	4620
acgcgcgaga	cgaaagggcc	tcgtgatacg	cctattttta	taggttaatg	tcataataat	4680
aatgggtttct	tagacgtcag	gtggcacttt	tcggggaaat	gtgcgcggaa	cccctatttg	4740
tttatttttc	taaataatgt	tccgctcatg	agacaataac	agacaataac	cctgataaat	4800
gcttcaataa	tattgaaaaa	ggaagagtat	gagtattcaa	catttccgtg	tcgcccttat	4860
tccctttttt	gcggcatttt	gccttctcgt	ttttgctcac	ccagaaacgc	tggtgaaagt	4920
aaaagatgct	gaagatcagt	tgggtgcacg	agtgggttac	atcgaactgg	atctgaatccg	4980
ggtaagatc	cttgagagtt	ttcgcccga	agaacgtttt	ccaatgatga	gcacttttaa	5040
agttctgcta	tgtggcgcg	tattatcccg	tattgacgcc	gggcaagagc	aaactcgtcg	5100
ccgcatacac	tattctcaga	atgacttggt	tgagtactca	ccagtcacag	aaaagcatct	5160
tacggatggc	atgacagtaa	gagaattatg	cagtgtgcc	ataaccatga	gtgataaacac	5220
tgccggccaac	ttacttctga	caacgatcgg	aggaccgaag	gagctaaccg	cttttttgca	5280
caacatgggg	gatcatgtaa	ctcgccctga	tcgttgggaa	ccggagctga	atgaagccat	5340
accaaacgac	gagcgtgaca	ccacgatgcc	tgtagcaatg	gcaacaacgt	tgcgcaaact	5400
attaactggc	gaactactta	ctctagcttc	ccggcaacaa	ttaatagact	ggatggaggc	5460
ggataaaagt	gcaggaccac	ttctgcgctc	ggcccttccg	gctggctggg	ttattgctga	5520
taaactctgga	gcgggtgagc	gtgggtctcg	cggatcatt	gcagcactgg	ggccagatgg	5580
taagccctcc	cgtatcgtag	ttatctacac	gacggggagt	caggcaacta	tggtgaacg	5640
aaatagacag	atcgctgaga	taggtgcctc	actgattaag	cattggtaac	tgtcagacca	5700
agttttactca	tatatacttt	agattgatth	aaaacttcat	ttttaattta	aaaggatcta	5760
ggtgaagatc	ctttttgata	atctcatgac	caaaatccct	taacgtgagt	tttcgttcca	5820
ctgagcgtca	gaccccgtag	aaaagatcaa	aggatcttct	tgagatcctt	tttttctgcg	5880
cgtaatctgc	tgcttgcaaa	caaaaaaacc	accgctacca	gcgggtgggtt	gtttgccgga	5940
tcaagagcta	ccaactcttt	ttccgaagg	aactggcttc	agcagagcgc	agataccaaa	6000
tactgtcctt	ctagtgtagc	cgtagttagg	ccaccacttc	aagaactctg	tagcaccgcc	6060
tacatacctc	gctctgctaa	tcctgttacc	agtggctgct	gccagtggcg	ataagtctgt	6120
tcttaccggg	ttggactcaa	gacgatagtt	accggataag	gcgcagcggg	cgggctgaac	6180
gggggggttcg	tgcacacagc	ccagcttgga	gcgaacgacc	tacaccgaac	tgagatacct	6240
acagcgtgag	ctatgagaaa	gcgccacgct	tccgaagggg	agaaaaggcgg	acaggatcc	6300
ggtaagcggc	agggctcgaa	caggagagcg	cacgagggag	cttccagggg	gaaacgcctg	6360
gtatctttat	agtccctgtc	ggtttcgcca	cctctgactt	gagcgtcgat	ttttgtgatg	6420
ctcgtcaggg	ggcggagcc	tatggaaaaa	cgccagcaac	gcggcctttt	tacggttcct	6480
ggccttttgc	tggecttttg	ctcacatggc	tcgacagatc	t		6521

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/17452

A. CLASSIFICATION OF SUBJECT MATTER		
IPC(7) : C07H 21/04; C12N 15/85,87,90; A01N 43/34; C12N 15/09		
US CL : 536/23.1; 435/320.1,325,419,455,467; 514/44; 800/21		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) U.S. : 536/23.1; 435/320.1,325,419,455,467; 514/44; 800/21		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) APS (EAST); STN (MEDLINE, BIOSIS, CAPLUS)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 01/07572 A2 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 01 February 2001 (01.02.2001), see entire document.	1,2,7,8,11,23- 25,27,50
X --- Y	WO 00/11155 A1 (THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY) 02 March 2000 (02.03.2000), see entire document.	3-6,9,10,12,13,26,28- 1,2,7,8,11,23- 25,27,50
X --- Y	US 6,171,861 B1 (HARTLEY et al.) 09 January 2001 (09.01.2001), see entire document, especially sequence listing.	3-6,9,10,12,13,26,28- 1-3,8,10-12,23- 27,33,36,39,40,43- 45,50,54,55,85
		4-7,9,13-22,28- 30,32,34,35,37,38,41, 42,46-49,51-53,56,67- 71,84,86
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents:		
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention	
"E" earlier application or patent published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone	
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art	
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family	
"P" document published prior to the international filing date but later than the priority date claimed		
Date of the actual completion of the international search 01 November 2002 (01.11.2002)	Date of mailing of the international search report 27 NOV 2002	
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Daniel M Sullivan <i>Della Collier for</i> Telephone No. 703-308-0196	

Form PCT/ISA/210 (second sheet) (July 1998)